Essays in Public Finance and Labor Economics

Dissertation
for the Faculty of Economics, Business Administration and Information Technology of the University of Zurich

to achieve the title of
Doctor of Philosophy
in Economics

presented by

Sebastian Findeisen
from Germany

approved in September 2013 at the request of

Prof. Dr. Fabrizio Zilibotti
Prof. Dr. Josef Zweimüller

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorises the printing of this Doctoral Thesis, without thereby giving any opinion on the views contained therein.

Zurich, September 18, 2013

Chairman of the Doctoral Comittee: Prof. Dr. Dieter Pfaff

Dissertation Abstract and Overview

The first chapter, jointly written with Dominik Sachs, studies the optimal design of integrated education finance and tax systems. The distribution of wages is endogenously determined by the costly education decisions of heterogeneous individuals before labor market entry. Consistent with empirical evidence, this human capital investment decision is risky. We find that an integrated education and tax system in which the government provides education loans to young individuals coupled with income-contingent repayment can always be designed in a Pareto optimal way. We present a simple empirically driven application of the framework to US data in which individuals make a college entry decision. We find the optimal repayment schemes for college loans can be well approximated by a schedule that is linearly increasing in income up to a threshold and constant afterwards. So although the full optimum could lead to complicated non-linear schedules in theory, very simple instruments can replicate it fairly well. The welfare gains from income-contingent repayment are significant.

In Chapter two, co-authored with Wolfgang Dauth and Jens Suedekum, we analyze the effects of the unprecedented rise in trade between Germany and "the East" – China and Eastern Europe – in the period 1988–2008 on German local labor markets. Using detailed administrative data, we exploit the cross-regional variation in initial industry structures and use trade flows of other high-income countries as instruments for regional import and export exposure. We find that the rise of "the East" in the world economy caused substantial job losses in German regions specialized in import-competing industries, both in manufacturing and beyond. Regions specialized in export-oriented industries, however, experienced even stronger employment gains and lower unemployment. In the aggregate, we estimate that this trade integration has caused some 493,000 additional jobs in the economy and contributed to retaining the manufacturing sector in Germany. We also conduct our analysis at the individual worker level, and find that trade had a stabilizing overall effect on employment relationships.

In Chapter three, co-authored with Dominik Sachs, we study optimal nonlinear taxation of labor income and linear taxation of capital income in a life cycle framework. Consistent with what empirical papers have found, inequality changes over the life cycle, both because of forecastable heterogeneity across individuals and because of idiosyncratic risk. Consistent with current tax practices, taxes condition on current (annual) earnings. We show that a simple formula for the optimal taxation of capital arises in our life cycle model. It follows a standard public finance trade-off: the tax rate tends to increase with wealth inequality and to decrease with how elastic savings are with respect to capital taxes. The tax rate tends to be positive if marginal social welfare weights are negatively correlated with wealth. In numerical simulations, we confirm the intuition from the theoretical analysis of the model that the government taxes capital income at positive rates. If capital taxes are allowed to be age-dependent, they increase over the life cycle, as wealth concentration increases. If labor taxes are allowed to be age-dependent, they also increase over the life cycle, as labor income inequality increases.

Finally, the last chapter, jointly written with Dominik Sachs, studies the implications of limited commitment on education and tax policies chosen by benevolent governments. Individual wages are determined by both innate abilities and education levels. Consistent with real world practices, the government can decide to subsidize different levels of education at different rates. Deviations from full commitment tend to make education policies more progressive, increasing the education subsidy for initially low skilled agents and decreasing it for initially high skilled agents. We provide suggestive cross-country correlations for this mechanism.

Acknowledgements

I am deeply indebted to my advisor Fabrizio Zilibotti who has provided me with strong support, guidance and encouragement from my very first day at the University of Zurich, throughout all stages of my dissertation and the job market process. I have constantly profited from his wisdom in the last four years. He has also served as a superb role model. It was inspiring to witness his enthusiasm and commitment towards economic research. I can only hope that some of it has rubbed off on me.

I am grateful to Kjetil Storesletten who has provided me with outstanding support on the academic job market, despite being affiliated with a different institution. I have benefitted from many discussions and support from Filippo Brutti, Dominic Rohner and especially Christoph Winter. In addition, I thank Christoph for his great support on the job market. I thank Josef Zweimüller for acting as a second reader for this thesis.

It has been an enormous privilege to co-author the Public Finance chapters of this thesis with Dominik Sachs. Our interactions in the last three years were a constant source of learning for me – and they still continue to do so. I thank Dominik for his relentless commitment to our common projects in the past and presence. I also feel grateful for his moral support as a friend throughout my graduate studies. It has been a pleasure to work with Wolfgang Dauth and Jens Südekum on the second chapter of the thesis. I am additionally grateful to Jens, for having served as my first mentor, introducing me to research and sparking my interest in economics when I was still an undergraduate at the University of Konstanz.

I would like to thank my fellow graduate students and friends in Zurich for cooperation, friendship and discussions about research, especially Andreas Müller, Peter Rosenkranz, David Schönholzer, Andreas Steinhauer and Yikai Wang. During the very last stage of my dissertation, I have been visiting the Department of Economics at UC Berkeley. I am indebted to Emmanuel Saez for his mentoring, comments on the third chapter and hospitality during my stay.

I would like to express my gratitude towards my parents Pauliana and Peter for their unconditional support throughout my whole life and everything they did for my personal and professional development. Finally, the deepest of my gratitude goes to Stefanie for unselfishly standing by my side in all the years and fully supporting every single step of mine. I am not sure if I had made it to this point without her. I am sure that it would have been much harder and less fulfilling without her kindness and presence in my life.

Contents

Chapter 1

Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans[1]

This chapter is joint work with Dominik Sachs. It is a revised version of Working Paper No. 40, *Working Paper Series*, Department of Economics, University of Zurich.

It is submitted for publication to the *Journal of Public Economics*.

1.1 Introduction

How should governments design their higher education finance systems? There exist large differences across countries in the structure of higher education finance. In some countries, such as Denmark, Finland and Sweden, university and college students

[1]We are grateful to our advisors Friedrich Breyer and Fabrizio Zilibotti for ongoing support and valuable comments. We also thank Manuel Amador, Dan Anderberg, Carlos da Costa, Emmanuel Farhi, Mike Golosov, Bas Jacobs, Sebastian Koehne, Normann Lorenz, Elena Mattana, Emmanuel Saez, Florian Scheuer, Dirk Schindler, Kjetil Storesletten, Aleh Tsyvinski, Matthew Weinzierl, Iván Werning, Christoph Winter and many seminar audiences for helpful discussions. We thank Stefan Voigt for valuable research assistance. We are grateful to Yale and Stanford for their hospitality, where parts of this paper were written. Sebastian Findeisen acknowledges the support of the University of Zurich (Forschungskredit of the University of Zurich, grant no. 53210603).

pay low or no tuition fees and in addition receive grants because of generous public subsidies for higher education. These countries have highly progressive tax systems, which allow to finance these education subsidies. By contrast, in the United Kingdom and the United States, e.g., the burden of educational costs mainly lies on the student and higher education is much less heavily subsidized by public finances. Instead, student loans offered by both the private and public sector play a big part in financing higher education. From a policy perspective, the choice of an optimal education finance system is intimately linked to the tax system. Both underlie the same basic trade-offs, namely equity concerns in the form of redistribution and insurance against income risk versus efficiency concerns by distorting labor supply and education incentives.

In this paper, we address the optimal design of integrated education finance and tax systems. We build a novel optimal taxation framework in the spirit of Mirrlees (1971) and the vast literature following his footsteps, which allows to study the question from a new angle. In our framework, the distribution of wages is not exogenous but determined by the costly education decisions of individuals before labor market entry. Consistent with what is typically found in empirical studies, this human capital investment decision is risky. To solve the problem we use an applied mechanism design approach. The benevolent government can observe total income and the education level of individuals, but it has to respect incentive compatibility – first, when individuals decide on education and second, when individuals decide on labor supply. The main novelty of our approach is that in our framework the government is not restricted to the use of predetermined instruments but is free to choose its own instruments, which can condition on education, income and savings. In addition, they are allowed to be fully non-linear.

We find that an integrated education and tax system in which the government provides education loans to young individuals coupled with income-contingent repayment rates of these loans after individuals enter the labor market can effectively deal with all the major trade-offs underlying the education finance and tax problem. In other words, such systems can always be designed such that they are second-best Pareto efficient. This is because income-contingent repayment rates allow the government to *effectively differentiate tax distortions across education groups*, minimizing the efficiency cost of labor supply distortions. At the same time, it can subsidize education by varying the generosity of the loans.[2] Importantly, in general the government typically will find it optimal that some individuals partially default and never pay back the full value of their loans, while for some individuals the amount of repayment might exceed their loan values because this provides insurance.

[2] We do not model credit market imperfection in the form of borrowing constraints. If these are relevant, as is still a debated question in the literature (Carneiro and Heckman 2005), wide availability of student loans has the additional benefit of lifting these constraints.

We present a simple empirically driven application of the framework to US data in which individuals make a college entry decision. We simulate optimal income taxes and college student loans with income-contingent repayment. The optimal policy simulation provides three important insights. First, we find the optimal repayment scheme for college loans can be well approximated by a schedule that is linearly increasing in income. So although the full optimum could lead to complicated non-linear schedules in theory, very simple instruments can replicate it fairly well. Second, for our benchmark parameterization college graduates find it optimal to participate voluntarily in the loan schemes as compared to taking a risk-free loan on the private market. Third, we calculate the welfare gains of moving from a third-best scenario where the government optimally sets the income tax and offers a loan system with non-contingent repayment to the system with contingent repayments. We find welfare gains ranging from about 0.2% to 0.6% of lifetime consumption and we show how these gains vary with risk-aversion.

Several countries like the United Kingdom, Australia and New Zealand currently administer income-contingent college student loans, where repayment is proportional to income.[3] Our framework gives these policies a theoretical second-best foundation, based on an applied mechanism design approach to the education finance and taxation problem. Our theoretical considerations suggest that it might be optimal for the government to enforce that very rich individuals pay back more than the capitalized loan value or that repayment might actually be decreasing in income. In the mentioned countries, repayment never exceeds the loan value and repayment schedules are non-decreasing in income. To address these issues, we also consider policy experiments in which we restrict income-contingent repayment not to exceed the actual loan value and to be non decreasing in income. We find that a large share of the welfare gains from the full optimum can be reaped with these simpler policies.

Relation To Existing Literature. Our paper makes a contribution to the literature on optimal income taxation starting from Mirrlees (1971) (see the recent survey of Piketty and Saez 2013). In Section 3 we discuss how the expression for optimal education-dependent marginal tax rates compares to the seminal optimal tax formulas from Diamond (1998) and Saez (2001) with exogenous human capital.

In two papers Bovenberg and Jacobs (2005) and Jacobs and Bovenberg (2011) analyze how endogenous education alters the optimal tax problem and discuss to what extent education should be subsidized. Bohacek and Kapicka (2008) study a dynamic model with certainty and obtain equivalent results regarding education subsidies. These articles work under certainty whereas we take idiosyncratic human capital risk into account.

[3]Chapman (2006) provides a survey for practices in those and other countries. To the best of our knowledge, the first economist to endorse the idea was Milton Friedman (1955). He envisioned repayment amounts to be proportional to income, i.e. a linearly increasing repayment schedule. Something we find as an optimal policy in our simulation for the most part of the income distribution.

Using analytical results and numerical illustrations, we discuss in detail in Section 3 how our findings for optimal education subsidies in a general risky environment relate to their findings. Importantly, with idiosyncratic education risk, the necessity of education dependent labor wedges and income-contingent loans arises, as intuitively they can be understood as providing an additional source of insurance. As we discuss in Section 2, when we review some stylized empirical facts, there is strong evidence that uncertainty about college returns is important and matters for human capital investment decisions.[4]

Two recent papers, Best and Kleven (2013) and Kapcika and Neira (2013), study how human capital acquisition at the working age influence the optimal taxation problem. We focus on a different part of the human capital accumulation process, namely education before labor market entry. Importantly, both papers reasonably assume that tax policies cannot directly condition on human capital acquired while working. In contrast, we allow the government to use information about education before labor market entry in the tax code, as is done in the real world in some countries in the form of income-contingent student loans. In addition, our focus is on education finance instead of only tax policies.

Concerning the implementation of history-dependent allocations, our paper is related to Golosov and Tsyvinski (2006) who consider an environment with absorbing disability shocks and present an implementation in which disability insurance conditions on asset testing. Also in the context of optimal taxation, Scheuer (2012) considers differential taxation of profits and labor income; in our case a comparable logic applies for an endogenous education instead of an occupational choice.

Finally, taking a quantitative approach and working in the Ramsey tradition with simpler but given policy instruments, Krueger and Ludwig (2013) solve for the optimal income tax and education subsides in a rich macro model.

This paper is organized as follows. Section 1.2 contains the basics of the model. In Section 1.3, we investigate dynamic incentive compatibility and describe the major properties of constrained efficient allocations. Decentralized implementations of constrained efficient allocations are provided in Section 1.4. We apply our model to the case of a binary education decision in Section 1.5 and Section 1.6 concludes.

[4]One strand of literature has looked at first- versus second-best investment rules of human capital under risk with a representative agent. Da Costa and Maestri (2007) show that human capital should always be encouraged in the second-best optimum. Anderberg (2009) emphasizes that the risk properties of human capital are crucial for the question whether and how education should be distorted relative to a first-best rule. Focusing on linear policy instruments, Anderberg and Andersson (2003) as well as Jacobs, Schindler and Yang (2012) obtain similar results. An early treatment how taxes affect the risk properties of human capital investment is Eaton and Rosen (1980). Grochulski and Piskorski (2010) focus on the implications of unobservable human capital investment for capital taxation in an ex-ante homogeneous agent setting with uncertainty. Kapicka (2006) introduces non-observable endogenous human capital into a dynamic, non-stochastic Mirrlees model where taxes can only be conditioned on current income. He shows that marginal tax rates are lowered due to the education margin.

1.2 The Model

1.2.1 Structure

We consider a stripped-down life-cycle model, in which individuals acquire formal education early in their life cycle and work afterwards. Individuals differ in innate ability θ, which can be interpreted as a one dimensional aggregate of (non-)cognitive skills, I.Q. and family background, and is distributed in the interval $[\underline{\theta}, \overline{\theta}]$ according to the cumulative density function (cdf) $F(\theta)$. After individuals learn their type θ, which is private information, they make a monetary educational investment e. Flow utility during education is denoted by $u^e(c_e)$ with $u^e_c > 0, u^e_{cc} < 0$. It takes T_e periods (years) until education is finished; the yearly education costs are denoted by e. To simplify exposition, we assume all levels of education take the same amount of time, an assumption we relax later in our optimal policy simulations

As individuals enter the labor market, they draw their labor market ability a from a continuous conditional cdf $G(a|e,\theta)$, which depends on *innate* ability θ and education e and has bounded support $[\underline{a}, \overline{a}]$. We assume that preferences over consumption and leisure are given by the utility function $u^w(c_w, l)$, where labor effort l is equal $\frac{y}{a}$. We assume that $u^w(\cdot,\cdot)$ obeys the Spence-Mirrlees condition. The working life lasts for T_w periods.

Expected lifetime utility of an individual of type θ is hence given by

$$\sum_{t=1}^{T_e} \beta^{t-1} u^e(c_e(\theta)) + \int_{\underline{a}}^{\overline{a}} \sum_{t=T_e+1}^{T_e+T_w} \beta^{t-1} u^w \left(c_w(\theta,a), \frac{y(\theta,a)}{a} \right) dG(a|\theta, e(\theta)), \tag{1.1}$$

where we assume the allocation within the education and working period to be constant.[5] We will write $\beta^e = \sum_{t=1}^{T_e} \beta^{t-1}$ and $\beta^w = \sum_{t=T_e+1}^{T_e+T_w} \beta^{t-1}$.

As equation (1.1) reveals, we abstract from further shocks to idiosyncratic labor productivity once individuals have entered the labor market. This simplifies and helps to focus the analysis on the education-taxation link. In the empirical literature, there is no ultimate consensus on the relative importance of heterogeneity before labor market entry (versus the role of shocks over the working life) for lifetime inequality, but different approaches have attributed a major role to it.[6]

[5] This is akin to the assumption that the first-order conditions of the second-best problem we solve are also sufficient.

[6] In recent work, Huggett, Ventura and Yaron (2011) estimate a structural life-cycle model and find that differences realized at the age of 23 can account for more of the variation in lifetime outcomes than do shocks received over the working lifetime. A standard reference is Kean and Wolpin (1997), who attribute 90% to heterogeneity realized before labor market entry, while Storesletten, Telmer and Yaron (2004) estimate a number of about 50%.

Nevertheless, we capture many empirical regularities with this specification of the model. First, assuming $G(a|e,\theta)$ to be non-degenerate, our model captures the important fact of uncertainty in the labor market and risky educational investment. See e.g. Cunha and Heckman (2008) or Chen (2008) for recent contributions.

Second, we allow this cdf to be a function of innate ability θ and thereby capture the fact that inequality in earnings is – to a certain extent – also determined by innate ability. Taber (2001) and Hendricks and Schoellman (2012a) suggest that much of the rise in the college premium may be attributed to a rise in the demand for unobserved skills, which are predetermined and independent of education. Indirect evidence for the importance of unobserved skills comes from the strong persistence of within-education-group inequality Acemoglu and Autor (2011).

Third, the cdf G being a function of e captures the returns to education. Importantly, for most of our results, we do not impose a certain assumption on the pattern of these returns.

Fourth, as long as $\frac{\partial^2 G(a|e,\theta)}{\partial\theta\partial z} \neq 0$, returns to educational investment differ in innate ability θ. E.g., Carneiro and Heckman (2005) document that the returns can differ by as much as 19% points across individuals for one year of college.[7]

To sharpen a few analytical results, it turns out helpful to place some structure on the behavior of $G(a|e,\theta)$:

Assumption 1: $G(a|e',\theta) \succeq_{FOSD} G(a|e,\theta) \Leftrightarrow G(a|e',\theta) \leq G(a|e,\theta)$, for all $e < e'$ and for all (θ, a).

Assumption 2: $G(a|e,\theta') \succeq_{FOSD} G(a|e,\theta) \Leftrightarrow G(a|e,\theta') \leq G(a|e,\theta)$, for all $\theta < \theta'$ and for all (e, a).

Assumption 3: $\frac{\partial^2 G(a|e,\theta)}{\partial\theta\partial e} \leq 0$ for all (θ, e, a).

These assumptions will not be needed to derive our main results, but help to illustrate important aspects of the model. Whenever an assumption is needed for a result, we refer to it. The first and the second one capture the notion that education and innate ability should both have a direct effect on labor market outcomes represented by a first-order stochastic dominance shift; a rather natural way of ordering distributions. The third one captures their interaction and respects the compelling evidence of complementarity between early ability and educational investment.

1.2.2 Definition of Wedges

For later purposes when we analyze optimal allocations and the respective tax and education finance systems that can implement such allocations, it is useful to define

[7] See also Lemieux (2006) for evidence on heterogeneity in returns.

wedges. They are equal to implicit marginal tax rates. We are particularly interested in labor and education wedges. We use subscripts to indicate partial derivatives.

Labor wedge: The labor wedge is positive (negative) if an individual works less (more) than it would at the intervention-free market price (which is her productivity level a). Formally the labor wedge reads as:

$$\tau_y(\theta, a) = 1 - \frac{u_l^w\left(c_w(\theta, a), \frac{y(\theta,a)}{a}\right)\frac{1}{a}}{u_c^w\left(c_w(\theta, a), \frac{y(\theta,a)}{a}\right)}.$$

Education wedge: Here, a positive (negative) wedge corresponds to an upward (downward) distortion of the education decision. Formally the education wedge reads as

$$\tau^e(\theta) = 1 - \frac{\beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta, a)\frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)}da}{\beta^e u_c^e(c_e(\theta))},$$

where $v^w(\theta, a)$ is the value function for the working period.

Finally, we will also look at optimal distortions of an individual's Euler equation between the education and the working period.

Savings wedge:

$$\tau_s(\theta) = 1 - \frac{\beta^e u_c^e(c_e(\theta))}{\beta^w R \int_{\underline{a}}^{\overline{a}} u_c^w\left(c_w(\theta, a), \frac{y(\theta,a)}{a}\right) g(a|e, \theta)da}$$

where R is the gross return on savings between the education and the working life. $\tau_s(\theta) > (<)0$ implies a downward (upward) distortion of savings.

1.3 Constrained Pareto Optimal Allocations

In this section, we characterize constrained Pareto efficient allocations, where 'constrained' refers to the government being unable to observe agents' type θ at the education stage and a in the working stage. In Subsection 1.3.1, we show that the problem is tractable using a first-order approach. In addition, we provide necessary as well as sufficient conditions for this approach to be valid. In Subsection 1.3.2, we analyze optimality conditions and their consequences for optimal policies. In Subsection 1.3.3, we explore the model using numerical simulations.

1.3.1 Incentive Compatibility

We cast the problem as a sequential mechanism – agents report an initial type θ in the education period and, after uncertainty has materialized, report their productivity a in the working period. The planner assigns initial consumption levels $c_e(\theta)$ and education levels $e(\theta)$ to individuals with innate ability θ. Moreover, with each report there comes a sequence of utility promises for the next period $\{v^w(\theta, a)\}_{a\in[\underline{a},\overline{a}]}$. In the second period, the screening takes place over consumption levels $c_w(\theta, a)$ and labor supply $y(\theta, a)$. All these quantities define an allocation in the economy. Dynamic incentive compatibility is ensured backwards, so we start analyzing the problem from the second period.

Education Period Incentive Compatibility

By the revelation principle, we can restrict attention to direct mechanisms. Suppose that in the first period agents have made truthful reports $r_\theta(\theta) = \theta$, albeit this is not necessary and just simplifies the exposition.[8] Conditions for this to be true are given in the next subsection. Conditional on this report, the second period incentive constraint must be met for any history of types (θ, a) and reporting strategy $r_a(a)$:

$$u^w\left(c_w(\theta, a), \frac{y(\theta, a)}{a}\right) \geq u^w\left(c_w(\theta, r_a(a)), \frac{y(\theta, r_a(a))}{a}\right) \qquad \forall a, r_a(a), \theta.$$

Like in a standard Mirrleesian problem preferences satisfy single-crossing for given first-period reports. For global incentive compatibility it is, hence, necessary and sufficient that all local envelope conditions hold:

$$\frac{\partial v^w(\theta, a)}{\partial a} = u_l^w\left(c_w(\theta, a), \frac{y(\theta, a)}{a}\right)\frac{y(\theta, a)}{a^2} \tag{1.2}$$

and the usual monotonicity condition, stating that $y(\theta, a)$ is non-decreasing in ability levels a, is satisfied:

$$\frac{\partial y(\theta, a)}{\partial a} \geq 0. \tag{1.3}$$

[8]The reason is that in the second period the utility is a function of $a, r_a(a)$ and $r_\theta(\theta)$ but not of θ.

Education Period Incentive Compatibility

In the education period, an agent takes into account the effect of her report about θ on future utility. Education period incentive compatibility is ensured if and only if the following double continuum of weak inequalities holds:

$$\beta^e u^e\left(c_e\left(\theta\right)\right) + \beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta, a) dG(a|e(\theta), \theta) \geq$$
$$\beta^e u^e\left(c_e\left(r_\theta(\theta)\right)\right) + \beta^w \int_{\underline{a}}^{\overline{a}} v^w(r_\theta(\theta), a) dG(a|e(r_\theta(\theta)), \theta) \; \forall \theta, r_\theta(\theta).$$

Let $V(\theta)$ be the associated value function. By using the FOC of an agent's reporting problem, one can easily derive the following envelope condition

$$\frac{dV(\theta)}{d\theta} = \beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta, a) \frac{\partial g(a|e(\theta), \theta)}{\partial \theta} da. \tag{1.4}$$

As often done in screening problems, our strategy for solving the second-best problem is to work with a relaxed problem with only restrictions (1.2) and (1.4) being imposed and then check ex-post whether incentive compatibility is fulfilled. In the numerical explorations in Section 1.3.3 we find that incentive compatibility is always satisfied and therefore the first-order approach is valid for the primitives we consider.[9]

Next, we present a set of sufficient conditions.

Lemma 1.3.1. *Suppose Assumptions 2 and 3 hold, conditions* (1.2), (1.3), (1.4) *are satisfied and we have:*
$(i) \frac{\partial y(\theta,a)}{\partial \theta} > 0$,
$(ii) \frac{\partial e(\theta)}{\partial \theta} > 0$,
then the considered allocation is incentive compatible.

Proof. See Appendix 1.7.1. □

This lemma implies that instead of directly ex-post verifying whether period one incentive compatibility is satisfied in an allocation, one can alternatively check these two simple monotonicity conditions; if they are fulfilled, then the allocation is incentive compatible. Whereas condition (ii) is always fulfilled in our numerical examples, condition (i) was often violated for very low a; we will comment on the reasons in Section 1.3.3 when we present numerical illustrations of the model.

[9]Our results of this section on dynamic incentive compatibility are related to previous work in the optimal non-linear pricing literature by Courty and Li (2000). They study optimal pricing schemes of a monopolist facing consumers with stochastic tastes. In our case the distribution of types tomorrow is endogenous since education is a choice. In recent contributions, Kapicka (2013) as well as Pavan, Segal and Toikka (2012) investigate the robustness and validity of the Mirrleesian first-order approach in a large class of general dynamic environments.

1.3.2 Properties of Constrained Pareto Optimal Allocations

The planner maximizes

$$\beta^e \int_{\underline{\theta}}^{\overline{\theta}} u^e(c_e(\theta))d\tilde{F}(\theta) + \beta^w \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} v^w(\theta, a)dG(a|e(\theta), \theta)d\tilde{F}(\theta)$$

subject to (1.2), (1.4) and the resource constraint:

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[\beta^e \left(c_e(\theta) + e(\theta)\right) + \beta^w \int_{\underline{a}}^{\overline{a}} (c_e(\theta, a) - y(\theta, a))dG(a|e(\theta), \theta)\right] dF(\theta) = 0.$$

We let the planner assign Pareto weights $\tilde{F}(\theta)$ to individuals, depending on their initial skill level. Any distribution of these weights, which we normalize to satisfy $\int_{\underline{\theta}}^{\overline{\theta}} \tilde{f}(\theta)d\theta = 1$, corresponds to one point on the Pareto frontier. λ_R denotes the multipliers on the resource constraint and $\eta(\theta)$ the multiplier function of the first-period envelope conditions. The planner uses the same discount rate as all individuals. We now characterize the wedges of second-best Pareto optimal allocations.

Labor Distortions

The following proposition characterizes the optimal labor wedge.[10] For expositional reasons, we focus on the case where utility is separable in consumption and labor and show the formula for the general case in the appendix.

Proposition 1.3.2. *Suppose preferences are separable of the form $u(c) - \Psi(l)$ where $u'' < 0$ and $\Psi'' > 0$ and further that $u(\cdot) = u^e(\cdot)$. At any constrained Pareto optimum, labor wedges satisfy:*

$$\frac{\tau_y(\theta, a)}{1 - \tau_y(\theta, a)} = \frac{1 + \varepsilon^u(\theta, a)}{\varepsilon^c(\theta, a)} \frac{u'(c_w(\theta, a))}{ag(a|e(\theta), \theta)} \left[\mathcal{A}(\theta, a) + \mathcal{B}(\theta, a)\right],$$

where

$$\begin{aligned}\mathcal{A}(\theta, a) =& G(a|e(\theta), \theta)\Bigg[\int_a^{\overline{a}} \frac{1}{u'(c_w(\theta, a^*))} dG(a^*|e(\theta), \theta) \\ & - \frac{1 - G(a|e(\theta)}{G(a|e(\theta), \theta)} \int_{\underline{a}}^{a} \frac{1}{u'(c_w(\theta, a^*))} dG(a^*|e(\theta), \theta)\Bigg]\end{aligned}$$

[10] In a recent paper Golosov, Troshkin and Tsyvinski (2011) provide formulas for dynamic optimal labor wedges with exogenous human capital, connecting them to empirical observables in the spirit of the contributions of Diamond (1998) and Saez (2001) for the static Mirrlees model.

$$\mathcal{B}(\theta, a) = \frac{1}{f(\theta)\lambda_R} \frac{\partial \left[1 - G(a|e(\theta), \theta)\right]}{\partial \theta} \eta(\theta),$$

where $\varepsilon^u(\theta, a)$ $(\varepsilon^c(\theta, a))$ *is the uncompensated (compensated) labor supply elasticity of type* (θ, a) *and*

$$\eta(\theta) = \tilde{F}(\theta) - \frac{\int_{\underline{\theta}}^{\theta} \frac{1}{u'(c_e(\theta))} f(\theta) d\theta}{\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{u'(c_e(\theta))} f(\theta) d\theta}.$$

Proof. See Appendix 1.7.2. □

To understand this analytical result and relate it to the literature, first assume that there would be no incentive problem in period 1, i.e. θ would be observable. In this case, the term $\mathcal{B}(\theta, a)$ would be zero everywhere because $\eta(\theta)$ would be zero everyhwhere. Then, the optimal labor wedge schedules for different values of θ would be the optimal insurance arrangement for the respective θ-type against income risk. In fact,we show that the formula resembles the standard formula of Saez (2001) for $\mathcal{B}(\theta, a) = 0$ in Appendix 1.7.2. If θ were observable, it would be an immutable *tag*. The planner would want to condition optimal insurance arrangements on θ in an Akerlof (1978) tagging manner.[11] The interpretation would be very standard that optimal effective marginal tax rates are decreasing in the compensated elasticity, typically larger for higher values of risk aversion and that the nonlinear shape of these effective marginal tax rates is to a large extent determined by the respective distribution function $G(a|\cdot, e(\cdot))$.

With θ being unobservable, the government has to take incentive compatibility in the first period into account. This is captured by term $\mathcal{B}(\theta, a)$. First, note that this term is proportional to the respective value of the Lagrangian-multiplier function $\eta(\theta)$. As long as the planner values the utility of low θ-types sufficiently high (i.e. such that $\tilde{F}(\theta)$ is not too low), $\eta(\theta)$ is positive. This is fulfilled for the Utilitarian case, but also for points on the Pareto frontier more in favor of higher θ-types. If, in addition, the rather natural Assumption 2 applies, term $\mathcal{B}(\theta, a)$ is unambiguously positive and thus is a force towards higher effective marginal labor income tax rates.

To get an intuitive understanding for this term, it helps to think about a stylized example. Assume that $G(a|e(\theta^*), \theta^*) = 1$ for all $a > a^*$. Thus, given their choice e, individuals of type θ^* have a zero probability of having a larger labor market skill than a^*. In contrast, assume that $G(a|e(\theta^*), \theta^* + \varepsilon) < 1$ for $\varepsilon \to 0$. In that case $\frac{\partial[1-G(a|e(\theta^*),\theta^*)]}{\partial \theta} \to \infty$ for all $a > a^*$ and therefore $\tau_y(\theta^*, a^*) = 1$ for all $a > a^*$. Intuitively, effective marginal tax rates of 100% for individuals of type (θ^*, a) with $a > a^*$ have no costs as the mass of

[11]More recently tagging is investigated by Cremer, Gahvari and Lozachmeur (2010), Mankiw and Weinzierl (2010) as well as Weinzierl (2012).

individuals whose behavior is distorted is equal to zero. At the same time, these high marginal tax rates make it less attractive for the type with ability $\theta^* + \varepsilon$ to mimick the θ^*-type.

In addition note that the education choice does not play a direct role for the results. If there were no education choices in the first period but individuals would just do nothing, Proposition 1.3.2 would be unchanged. Education only has an indirect choice on the value of optimal effective marginal tax rates through its impact on the distribution of skills. This does not imply that the planner does not take into account the adverse effects of labor supply distortions on the education margin. In fact, the social planner takes this into account by subsidizing the education margin as we discuss in the next subsection.

Finally, a no-distortion at the top and bottom result goes through since $\mathcal{B}(\theta, \overline{a}) = \mathcal{B}(\theta, \underline{a}) = \mathcal{A}(\theta, \overline{a}) = \mathcal{A}(\theta, \underline{a}) = 0$.

Education Distortions

The following proposition characterizes optimal education policies.

Proposition 1.3.3. *At any constrained Pareto optimum, the education wedge is given by:*

$$\begin{aligned}\tau^e(\theta) =& \frac{\beta^w}{\beta^e} \int_{\underline{a}}^{\overline{a}} (y(\theta, a) - c_w(\theta, a)) \frac{\partial g(a|e(\theta), \theta)}{\partial e(\theta)} da \\ &+ \frac{\beta^w}{\beta^e} \frac{\eta(\theta)}{\lambda_R f(\theta)} \int_{\underline{a}}^{\overline{a}} \frac{\partial v^w(\theta, a)}{\partial a} \frac{\partial^2 G(a|e(\theta), \theta)}{\partial e(\theta) \partial \theta} da.\end{aligned}$$

Proof. See Appendix 1.7.2. □

The first term captures the expected marginal fiscal gain of an increase in education. One can show that it is always positive under Assumption 1 (FOSD shift of education) and positive labor wedges. Investing a dollar more into education increases the expected obligation of an agent. The first part of the education wedge exactly offsets this effect from the labor wedge. Bovenberg and Jacobs (2005) have discovered this effect for the static Mirrlees model, whereas we show this fiscal externality part of the wedge extends to the setting with uncertainty, holding in expectation.

We now turn to the second term. Under Assumption 3 the cross-derivative $\frac{\partial^2 G(a|e(\theta),\theta)}{\partial e(\theta)\partial \theta}$ is negative and $\frac{\partial v^w(\theta,a)}{\partial a}$ is positive everywhere by second-period incentive compatibility. Further, as discussed in the previous subsection, $\eta(\theta)$ is positive along a large part of the Pareto frontier. Then the second part of the education wedge is negative and acts as an implicit tax on education. By distorting education downward, the planner relaxes binding incentive constraints and can redistribute more effectively in line with her preferences. This is a consequence of the complementarity assumption, stating that agents endowed with higher innate skills gain more from education at the margin. The bundle of a lower type, hence, becomes less attractive from the perspective of an agent if education is downward distorted. Such an intuition is familiar from the standard static Mirrlees model concerning positive marginal income tax rates on the interior of the skill set. Relatedly, for this incentive term a zero at the top and at the bottom $(\underline{\theta}, \overline{\theta})$ result holds.[12]

Savings Distortions

The characteristics of savings distortions depend on the properties of the utility function $u^w\left(c_w, \frac{y}{a}\right)$. In the case of separable preferences, the well explored inverse Euler equation holds (Diamond and Mirrlees (1978), Rogerson (1985), Golosov, Kocherlakota and Tsyvinski (2003)), making it optimal to tax savings at every initial skill level θ, improving the ability of the planner to provide labor supply incentives. In general, the sign of the wedge depends on the exact functional form assumption and especially on the interaction of labor effort and the marginal utility of consumption – see Golosov, Troshkin and Tsyvinski (2011) for an elaborate discussion of the underlying forces in a dynamic Mirrleesian model.

1.3.3 Numerical Illustration

In this section we numerically explore our model in an illustrative manner. We consider two skill distributions as our primitives $G(a|e, \theta)$ that lead to very similar equilibrium wage distributions and educational expenses for actual given policies. In one of the cases, the distribution function is characterized by a strong complementarity between innate skills and education. In the other, there is less complementarity and the direct effects of education and innate skills dominate.

We solve for the Utilitarian optimum, so $\tilde{f}(\theta) = f(\theta)\ \forall\ \theta$. The utility function is:

$$U(c, l) = \frac{c^{1-\rho}}{1-\rho} - \frac{(y/a)^{\sigma}}{\sigma},$$

where we set $\sigma = 3$, implying a Frisch elasticity of 0.5 and the CRRA coefficient to $\rho = 2$.

[12]Jacobs and Bovenberg (2011b) discuss deviations from a first-best rule for the education subsidy for a general earnings function in the case without uncertainty. Our result is similar to their first result that a complementarity in education and ability leads to a tax on education. They also consider the degree of complementarity between labor supply and education which might call for an education subsidy in contrast. This second effect disappears in our environment since the returns to labor supply – once uncertainty has materialized – are independent from the education choice.

We assume that labor market abilities are distributed log-normally following common practice and impose the location parameter μ to be a function of θ and e. Concerning θ, we assume a uniform distribution within $[0.1, 1]$.

Case (a) - Strong Complementarity: The functional form of the location parameter is:

$$\mu(\theta, e) = 1.7 + 1.5\theta^{0.5}e^{0.15}.$$

In this case, individuals are the same if they do not acquire any education at all. However, the more education they acquire the stronger are the differences in the location parameters. This inequality in μ will be reinforced by the fact that agents have incentives to self-select into different education levels because of heterogeneous returns.

Case (b) - Strong Direct Effect: In the second case we assume,

$$\mu(\theta, e) = 1.5 + \theta + 0.75e^{0.25}.$$

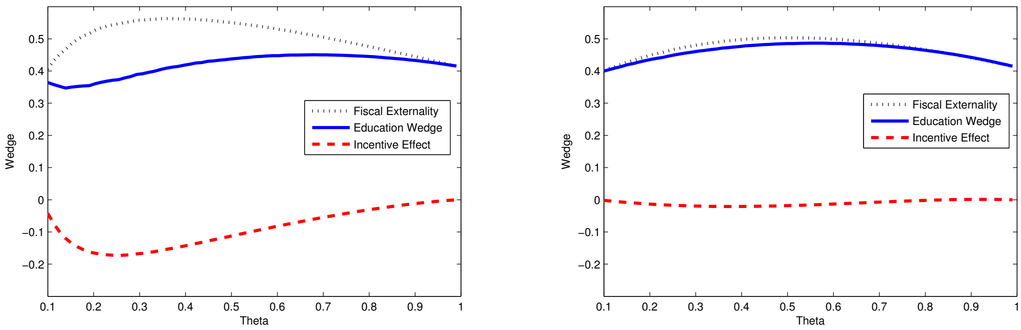
In this case, individuals are already very different from the outset, i.e. if nobody acquires any education. The difference in the location parameter then stays constant for a uniform increase in e across agents. Although $\frac{\partial^2 \mu(\theta,e)}{\partial\theta\partial e} = 0$ in this case, Assumption 3 is fulfilled for the relevant range. However, innate skills and education are weaker complements as compared to Case (a).

The respective parameters for the two cases as well as the respective constant marginal costs of education were chosen such that with an approximation of the current tax and college subsidy system in the US, the model roughly replicates per-capita expenditures on college education and the centers of the interval of the location parameters of the log-normal distributions is equal to the empirical value of the wage distribution.[13]

Figure 1.1 illustrates optimal education wedges for the two cases. In both cases, the optimal allocation features positive implicit education subsidies around 40%, which are relatively flat across innate types. The main difference between the two cases lies in the incentive effect. When innate skills and education are complements, the planner finds it optimal to tax education relative to a first best in line with Proposition 1.3.3. In Case (a) this incentive effect becomes as large as 17% whereas in Case (b) it hovers around zero.

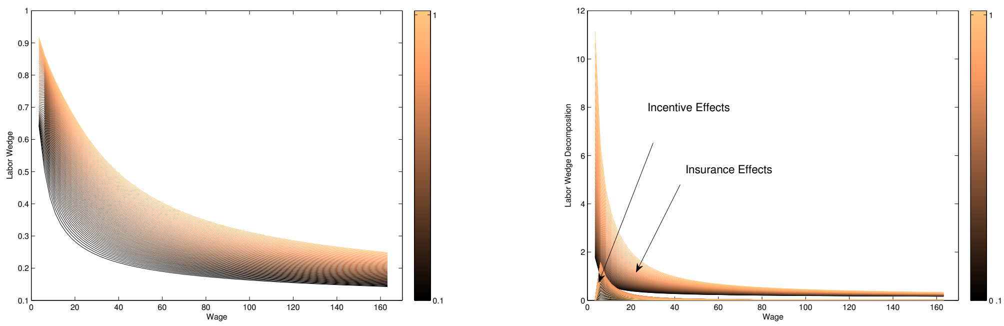
Figure 1.2 illustrates the optimal labor wedges from Proposition 1.3.2. Panel (a)

[13] Following Gallipoli, Meghir and Violante (2011) we set the labor income tax to a flat rate of 27% and a lump sum transfer of one sixth of labor income per capita. We introduce a yearly education subsidy of 24%. In both cases, for these given policy instruments, average college education expenditures per year are roughly 30% of yearly median income; a long run average for the US (Gallipoli, Meghir and Violante 2011). The realized values of $\mu(\theta, e)$ are within the range $[2.02, 3.34]$, centered around 2.76, the value of the lognormal fit for the US wage distribution found by Mankiw, Weinzierl and Yagan (2009); as them, we set the scale parameter equal to 0.565.



(a) Strong complementarity (b) Strong direct effect

Figure 1.1: Optimal Education Wedges



(a) Optimal Labor Wedges (b) Labor Wedge Decomposition

Figure 1.2: Optimal Labor Wedges

displays the optimal labor wedge as a function of income.[14] Darker regions refer to innate low types and lighter regions to innate high types. The picture shows that higher innate types face high labor wedges, wheres the shape of the wedges does not vary with θ.[15] In the next panel (b), we illustrate the decomposition from Proposition 1.3.2 into the insurance term and the incentive term by plotting $\mathcal{A}(\theta, a)$ and $\mathcal{B}(\theta, a)$. The set of insurance effects $\mathcal{A}(\theta, a)$ lies above the set of incentive effects $\mathcal{B}(\theta, a)$. Still, especially at the beginning of the income distribution incentive effects contribute to higher implicit tax rates. The graph also reveals that these incentive effects are of more importance for higher innate types on average.

[14]To economize on space we only show the figures for Case (b) here. The graphs for Case (a) turn out to look nearly identical.

[15]Since low incomes the distortions are strongly increasing in θ, condition (i) of Lemma 1.3.1 is typically not fulfilled for low a.

1.4 Implementation

So far we only considered a direct mechanism, in which individuals make reports about their realized type and the planner assigns bundles of consumption, labor supply and education as functions of the reports. The focus in the characterization of the optimal allocation was on wedges or *implicit* price distortions of the allocation. In this section, we explore two decentralized implementations of constrained Pareto optima. We focus on utility functions $u\left(c_w, \frac{y}{a}\right)$ with no income effects, i.e. $u\left(c_w - \Psi\left(\frac{y}{a}\right)\right)$ as in Diamond (1998) or Greenwood, Hercowitz and Huffman (1988). We do this for expositional purposes – in an earlier version of the paper (Findeisen and Sachs 2012) we discuss the implementation with income effects.[16] Additionally in the main body, we focus on implementations where education $e(\theta)$ is monotone in type and discuss the case where $e(\theta)$ may be non-monotone in Appendix 1.7.3; in this case policies are very similar.

1.4.1 Implementation One: Student Grants and Income Taxes Conditioning on Education

The benevolent government offers a menu of student grants to the agents. These grants $\mathcal{G}$ are conditional on education. In the working period, there is a tax function, which does not only condition on earnings but also on educational investment.

Proposition 1.4.1. *Suppose there are no income effects and education $e(\theta)$ is strictly monotone. Any constrained Pareto optimal allocation can be implemented by a grant schedule $\mathcal{G}(e)$, an education dependent income tax $T(y,e)$ and a savings tax $T^s(s)$, where*

- $\mathcal{G}(e(\theta)) = e(\theta) + c_e(\theta)$
- $T(y(\theta,a), e(\theta)) = y(\theta,a) - c_w(\theta,a)$
- $T^s(s)$ *is defined as in Appendix 1.7.3.*

Proof. See Appendix 1.7.3 □

Implementation of savings wedges: The savings function $T^s(s)$ is prohibitively high such that all agents choose s = 0, hence in this implementation there are no private savings. However, as shown in Werning (2011) this comes without loss of generality: by a Ricardian equivalence argument, we can adjust $G(e(\theta))$ and $T(y(\theta,a), e(\theta))$ with lump-sum transfers and deductibles to arrive with a non-linear savings tax schedule, which produces non-zero private savings for every agent and the same allocation with

[16] In general, there is also the issue of needing history dependent taxes even with constant wages over the working life, see Werning (2007). We are grateful to Bas Jacobs for pointing out that this can be overcome by the assumption of no income effects.

the same distortion of consumption across periods. The full argument can be found in Werning (2011).

Implementation of labor wedges: Agents enter the second period with no savings as argued above. Their budget constraint is then: $T(y(\theta, a), e(\theta)) = y(\theta, a) - c_w(\theta, a)$. From the agents' optimality conditions for y and c_w it follows that marginal tax rates $T_y(y(\theta, a), e(\theta))$ are equal to labor wedges $\tau_y(\theta, a)$ as characterized in Section 1.3.2.

Implementation of education wedges: In contrast to the optimal labor *wedge*, which equals the optimal labor *tax*, there is no single policy instrument for which the education wedge equals the marginal distortion of the policy. Instead, the government uses two instruments: i) the non-linear grant schedule $\mathcal{G}(e)$, which depends on education chosen and ii) the labor tax code in the second period. Using the agents' optimality conditions in the proposed implementation one can show that the wedge equals:

$$\tau^e(\theta) = \mathcal{G}'(e) - \int_{\underline{a}}^{\overline{a}} \frac{u'(c_w(\theta, a))}{u'(c_e(\theta))} g(a|e(\theta), \theta) T_e(y(\theta, a), e(\theta)) da$$

A positive value of $\tau^e(\theta)$ encourages education at level θ. The incentive for agents to increase their educational attainment comes from: i) An increase in their grant measured by $\mathcal{G}'(e)$[17] and ii) an increase or decrease in their labor income tax burden for all states, i.e. $T_e(y(\theta, a), e(\theta))$.

1.4.2 Implementation Two: Student Loans with Income-Contingent Repayment

The previous implementation required that people with the same income but different levels of education pay different taxes. In reality people might perceive this as a violation of horizontal equity concerns, which could hinder the political feasibility of such policies. In this light we now present a more appealing alternative implementation with only one labor income tax schedule and a repayment scheme of the education grant.[18] Technically, this can be seen as a simple reinterpretation of the previous implementation – we take the tax system of the $\underline{\theta}$-type as the *common* labor income tax schedule and introduce an income-contingent repayment schedule, which conditions on the size of the loan.[19]

[17]Theoretically it could be the case that $\mathcal{G}$ is (partly) decreasing in e if $c_e(\theta)$ is sufficiently decreasing. However, this is rather unlikely and in all our numerical examples we have $c_e'(\theta) > 0$.

[18]Diamond and Saez (2011) argue that practical policy prescription from optimal tax models should not go against commonly held normative views (horizontal equity for example) and limit complexity to a reasonable degree. The second implementation seems in line with these recommendations.

[19]Related implementations are of course possible where the tax function of another θ-type can be the labor income tax schedule in place. The extreme case would just be to say that income taxes do not exist and all schedules that were interpreted as history-dependent labor income schedules in implementation 1 can now be interpreted as repayment schedules. Taking the labor income tax schedule of the $\underline{\theta}$-type, however, seems to be more natural in our view. Especially in our application of the theory in Section 1.5.

Together both instruments are sufficient to replicate the optimal labor wedges. Formally we summarize this in the following proposition:

Proposition 1.4.2. *Suppose there are no income effects and education $e(\theta)$ is strictly monotone. Any constrained Pareto optimal allocation can be implemented by a (compulsory) loan schedule $L(e)$, a loan repayment schedule $\Gamma(y, L)$, an income tax $T(y)$ and a savings tax $T^s(s)$ where*

- $L(e(\theta)) = e(\theta) + c_e(\theta)$
- $\Gamma(y(\theta, a), L(e(\theta))) = c_w(\underline{\theta}, \tilde{a}(\underline{\theta}, y(\theta, a))) - c_w(\theta, a)$ *if* $y \in [y(\underline{\theta}, \underline{a}), y(\underline{\theta}, \overline{a})]$ *and* $\Gamma(y(\theta, a), L(e(\theta))) = y(\theta, a) - c_w(\theta, a)$ *otherwise.*
- $T(y) = y - c_w(\underline{\theta}, \tilde{a}(\underline{\theta}, y)) \; \forall \; y \; \in \; [y(\underline{\theta}, \underline{a}), y(\underline{\theta}, \overline{a})]$ *and* $T = 0$ *otherwise*
- $T^s(s)$ *is defined as in Appendix 1.7.3.*

where $\tilde{a}(\theta, y)$ *is the inverse of* $y(\theta, \cdot)$ *for a.*

Proof. The budget constraint of an individual reads as:

$$\begin{aligned} c_e(\theta) + e(\theta) &\leq L(e(\theta)) \\ c_w(\theta, a) &\leq y(\theta, a) - T(y(\theta, a)) - \Gamma(y(\theta, a), L(e(\theta))), \end{aligned}$$

which is equivalent to the budget constraint in Implementation 1 since by definition $\mathcal{G}(e) = L(e) \; \forall \; z$ and $T(y, z) = T(y) + \Gamma(y, z) \; \forall \; y, z$. Hence it is a direct consequence of Proposition 1.4.1. □

The similarity to the other implementation is apparent. Using the agents' optimality conditions, one can show that the education wedge equals

$$\tau^e(\theta) = L'(e) - \int_{\underline{a}}^{\overline{a}} \frac{u'(c_w(\theta, a))}{u'(c_e(\theta))} g(a|e(\theta), \theta) \Gamma_L(y(\theta, a), L(e(\theta))) \frac{dL(e(\theta)}{de} da,$$

and the labor wedge equals

$$\tau_y(\theta, a) = T'(y(\theta, a)) + \Gamma_y(y(\theta, a), L(e(\theta))).$$

Education wedges are implemented by the non-linear loans schedule and how repayment varies with education level. The labor wedge is equal to the marginal tax rate and how loan repayment varies with income.

Note that in Proposition 1.4.1, we assume the loans to be mandatory. In the numerical simulation we check whether this is a restrictive assumption by allowing college graduates to opt out and instead take a loan with fixed repayment, i.e. with a fixed interest rate. For our baseline parameterization, we find that college students participate voluntarily in the government loan system. Finally, notice that we did not impose a cap on repayments so that in theory for some income and education levels, they might exceed the capitalized loan values. In our numerical simulations, we also consider income-contingent repayment polices, which are not allowed to exceed the loans value.

1.5 An Application of the Model: College vs. High-School

We now present an empirically driven application of our model. We limit education to be a binary instead of a continuous choice. Agents either enter the labor market directly after high-school graduation or go to college before working. Additionally, we restrict the analysis to two levels of innate ability levels, one that refers to high school and one that refers to college. These simplifications enable us to parameterize the model using factual and, importantly, estimated counterfactual income distributions from the empirical labor literature (Cunha and Heckman 2007, 2008). Further, the simplification has the advantage that it is easy to incorporate foregone earnings as an implicit cost of education.

1.5.1 Parametrization

Individuals live for 47 years after they graduate from high-school (age 18-65). Afterwards they enter the labor market directly, or decide to go to college and graduate in four years. We label the two innate types θ_{HS} and θ_{CO}.[20] . The incentive constraints read as:

$$\begin{aligned} &\beta^e u(c_e) + \beta^{wCO} \int_{\underline{a}}^{\overline{a}} v_{CO}(a, \theta_{CO}) g(a|CO, \theta_{CO}) da \\ \geq &\beta^{wHS} \int_{\underline{a}}^{\overline{a}} v_{HS}(a, \theta_{HS}) g(a|HS, \theta_{CO}), \end{aligned} \tag{1.5}$$

and

$$\begin{aligned} &\beta^{wHS} \int_{\underline{a}}^{\overline{a}} v_{HS}(a, \theta_{HS}) g(a|HS, \theta_{HS}) da \\ \geq &\beta^e u(c_e) + \beta^{wCO} \int_{\underline{a}}^{\overline{a}} v_{CO}(a, \theta_{CO}) g(a|CO, \theta_{HS}) da, \end{aligned} \tag{1.6}$$

where $g(a|CO, \theta_{CO})$ and $g(a|HS, \theta_{HS})$ are the probability density functions (pdfs) of the factual ability distributions and $g(a|HS, \theta_{CO})$ and $g(a|CO, \theta_{HS})$ are the pdfs of the

[20] We assume that it is *a priori* optimal that the low type θ_{HS} chooses the lower educational attainment (high school) and that θ_{CO} chooses the higher educational attainment (college).

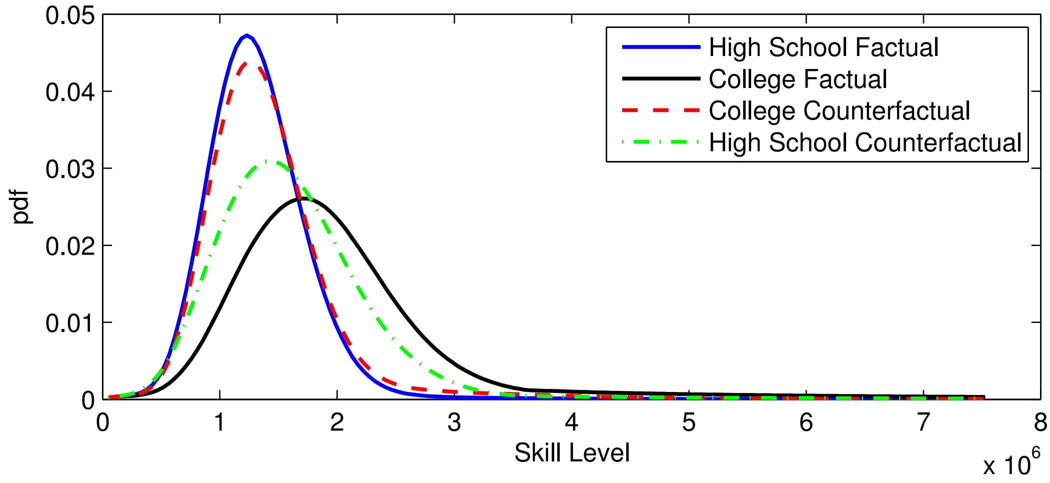


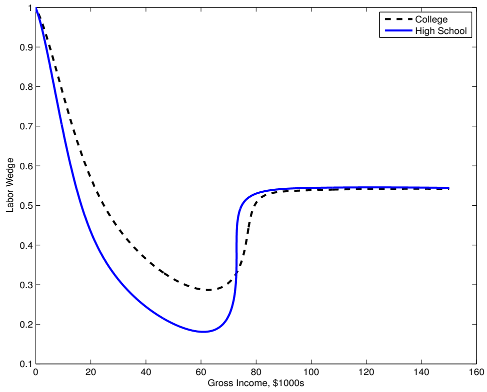
Figure 1.3: Skill Distributions

counterfactual ability distributions. The discount factors take into account the different lengths of the periods, i.e. $\beta^e = \sum_{t=1}^{4} \beta^{t-1}$, $\beta^{wCO} = \sum_{t=5}^{47} \beta^{t-1}$ and $\beta^{wHS} = \sum_{t=1}^{47} \beta^{t-1}$. Note that college types now have to be compensated for their foregone labor earnings, the implicit cost of college education. To get the ability distributions, we take the factual and counterfactual earnings distributions for high-school graduates plotted in Cunha and Heckman (2007) in Figures 1 and 2.[21] After using a kernel smoother, we append Pareto tails at earnings of $88,000. Finally, we smooth the resulting distribution again to overcome the kink from the appended tail. Given a (linear) approximation of the real world tax system we calibrate the implied skill distributions as input for the model from the optimality conditions of the agents.[22] We assume there is an atom of workers equal to five percent for each distribution reflecting unemployment or disability as in Mankiw, Weinzierl and Yagan (2009). The resulting calibrated skill distributions are illustrated in Figure 1.3. The share of high school and college types are set to 64.19% and 35.81%, respectively, as reported in Cunha and Heckman (2008). Following Abbott, Gallipoli, Meghir and Violante (2013), we set the annual monetary cost of college education to roughly a third of median income in our data. The yearly interest rate is set to 4% and the yearly discount factor β to $1/1.04$.. We work with a CRRA specification and focus on the case with no-income effects so that:

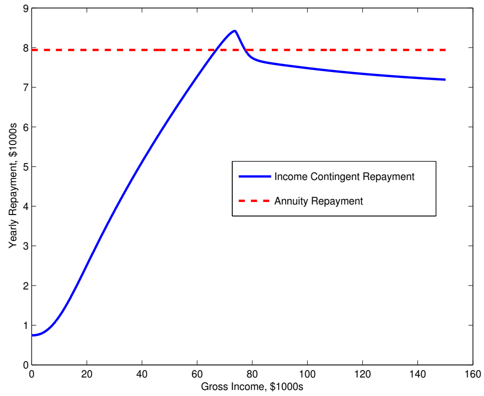
$$U(c, y, a) = \frac{\left(c - \frac{(y/a)^{\sigma}}{\sigma}\right)^{1-\rho}}{1-\rho},$$

[21]We used the software GetData Graph Digitizer to read out the data from the graphs. Since Cunha and Heckman (2007) consider the present value of lifetime earnings (18-65), we take a 47 years annuity with the same present value, i.e. we take something similar to average annual earnings.

[22]Saez (2001) has pioneered the approach to calibrate skill distributions from actual income distributions. We employ the same approximation as in the calibration of the full model in Section 1.3.3.



(a) Labor Wedges



(b) Repayment Schedule

Figure 1.4: Utilitarian Optimum

with $\sigma = 3$, implying a constant labor supply elasticity of 0.5 and set $\rho = 2$.[23]

1.5.2 Policies in Baseline Case

To the best of our knowledge, there exists no systematic evidence on the conditional distributions of top incomes for college graduates and non-graduates separately. In the baseline case, we conservatively assume identical tails for both groups, working with a Pareto parameter of 1.5 (Atkinson, Piketty and Saez 2011, Diamond and Saez 2011).

Second-Best Optimal Policies

Optimal Labor Wedges: Figure 1.4(a) displays the optimal labor wedges as a function of yearly income up to $160,000. Both schedules follow a U-shaped pattern, reflecting a result from the static Mirrlees problem (Diamond 1998, Saez 2001). The intuition for the pattern is simple: for very low incomes, marginal distortions are high for two reasons: first, distorting their labor supply is relatively harmless since they are rather unproductive. Second, the inverse hazard rate $\frac{1-G(a|\cdot,\cdot)}{g(a|\cdot,\cdot)}$ is rather high. Note that $1 - G(a|\cdot, \cdot)$ is proportional to the additional revenue generated by the (implicit or explicit) marginal tax rate and $g(a|\cdot, \cdot)$ is the mass of individuals whose labor supply is distorted. For intermediate incomes the density $g(a|\cdot, \cdot)$ strongly increases making distortions more and more harmful, leading to a decrease in optimal distortions. Finally,

[23] Note that savings are not distorted in our application. As we assume an education period of length zero for the high school type, there is no transition from an education to a working period, where the planner would find it optimal to distort savings for the high school type. For the college type, we get a no distortion at the top result for savings and therefore, for him the euler equation holds between the education and the working period.

due to the properties of the Pareto distribution, the ratio $\frac{1-G(a|\cdot,\cdot)}{ag(a|\cdot,\cdot)}$ converges to a constant and as a consequence the labor wedges start to converge.

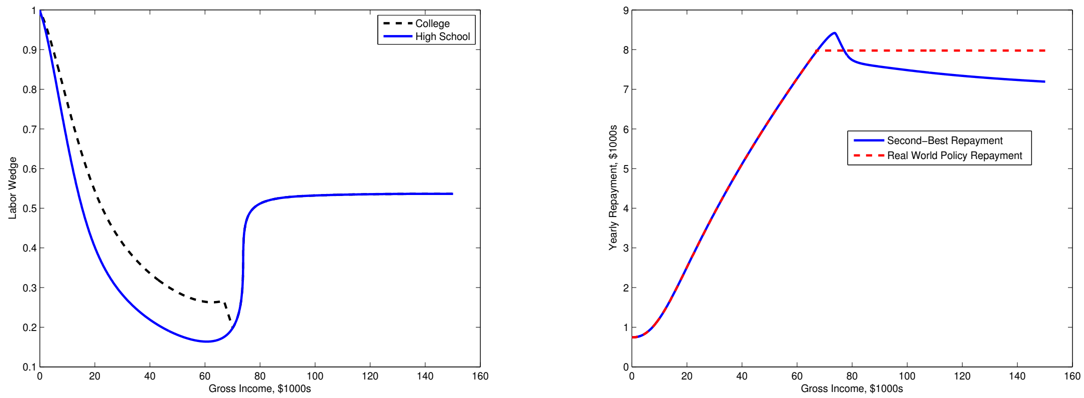
Looking at Figure 1.3, one can see in which way tax distortions are tailored to the different income distributions. At every point of the skill support before the Pareto tail kicks in, college labor distortions generate much bigger mechanical revenue effects for the government. In the top income tails, the wedges converge to almost the same top tax rate (Saez 2001),with a very small difference caused by the education incentive force $\mathcal{B}(\theta, a)$, which we discussed in the theoretical section of the paper, that leads to slightly higher top tax rates for high school types to increase the attractiveness of going to college.[24]

Repayment Schedule: We now build on the implementation results from the previous section and illustrate optimal income-contingent repayment schedules. The (common) labor income tax schedule is determined by the high-school labor wedges. Figure 1.4(b) shows the yearly repayment of college debt as a function of income. The slope of the repayment schedule is given by the difference in the labor wedges as we laid out in the previous section. As the college wedge lies above the high-school wedge, repayment is increasing in income up to incomes of US-$80,000. Repayments for college graduates start at about US-$1,000. Remarkably, the repayment schedule of loans is almost linear with a slope of roughly 0.1, because the difference in the labor wedge is almost constant. Afterwards, there is a very small decreasing range and the repayment schedule flattens out as the top labor wedges converge. In sum, optimal repayments can be very well approximated by an intercept of US-$1,000, a US-$1,000 increase in repayment for every US-$10,000 earned up to earnings of US-$70,000 and no additional repayments for incomes above that threshold. So although, we did not place any restrictions on the shape of the repayment schedule, linearity comes very close to the second-best optimum.

The red dotted horizontal line shows the yearly repayment that would occur if individuals chose a standard loan (with yearly interest of 4%) where the repayment is not contingent on income and they repay the same amount every year. As can be seen, only some individuals pay back more than that in the income-contingent case. This is sensitive to the interest rate, however. For 3%, e.g., more individuals would pay back more in the income-contingent case. For 5%, nobody would pay back more.

As discussed in the implementation section, we assumed the college loan system to be mandatory. We check if this is a restrictive assumption by allowing college graduates to opt out and instead take a loan with a yearly interest rate of 4% to finance tuition and early consumption. We find that given the choice, individuals would opt into the

[24]Some of these results are related to the simulations of Luttmer and Zeckhauser (2008) who consider a static setting where going to college is purely a signal and not an investment; hence counterfactual and factual distributions are equal.



(a) Labor Wedges

(b) Repayment Schedule

Figure 1.5: Real World Adjustment

loan system with income contingent repayment rates. This is also true for an interest rate of 3%. However, this is arguably a strict test of the assumption, since it is not clear if individuals would be able to borrow up to their desired amount and might face a substantial risk premium on their interest rate if they borrow in the private market.

Real World Policies: Cap on Repayment and Non-Decreasing Repayment

There might be two limitations to the full second-best optimum which could reduce its real world appeal. First, for some (small) range of the income distribution, repayment for college graduates actually exceeds the loan value, as obvious from Figure 1.4(b). Second, for high earners the repayment schedule actually decreases in income. These properties are likely to go against commonly held normative views, when it comes the actual implementation of an income-contingent loan system. Indeed, actual income-contingent repayment systems in the UK or Australia are never decreasing and cap repayment at the loans values. To deal with these concerns, we calculate an allocation which can be implemented with a repayment schedule respecting these constraints – i.e. it is never decreasing and capped at the loan value. In this scenario, effective marginal tax rates for college graduates are adjusted so that they are equal to the marginal tax rates for high school graduates as soon as repayment reaches the capitalized loan value. These modified polices still respect incentive compatibility and budget feasibility, of course.[25] Figures 1.5(a) and 1.5(b) show the resulting labor wedges and the repayment schedule. By construction, this repayment system is, of course, inferior in welfare terms

[25]More technically, we first adjusted the lump sum element of the common labor income tax schedule such that the government budget constraint holds. In case, the resulting allocation is not incentive compatible, we adjusted the lump sum elements of the labor income tax and the repayment schedule such that the government budget constraint holds and the incentive constraint of the college type binds.

to the optimal repayment schedule. As we show in the next subsection, this welfare loss is small.

The Welfare Gains From Income-Contingent Repayment

We now aim at quantifying what the potential welfare effects of income-contingency might be and how much of these welfare gains can be obtained by the (ad-hoc) adjusted repayment schedules, which respect a 'no-decreasing constraint' and put a cap on repayment.

The natural policy comparison is the case where repayment is not contingent on income. For this benchmark case, we allow the government to freely choose an income tax schedule and also optimize over education subsidies and savings taxes. Formally the only additional restriction is that individuals with the *same income* should face the *same labor wedge*.

To be able to make such a welfare comparison, the crucial assumption is the absence of income effects. In this case, the restriction that the labor wedge is only a function of current income is simply *equivalent* to:[26]

$$y(\theta, a) = y(a). \tag{1.7}$$

The following proposition states how a Pareto optimal allocation subject to (1.7) can be implemented in the binary education model.[27]

Proposition 1.5.1. *Assume there are no income effects. Then any Pareto optimal allocation subject to private information and (1.7) can be implemented by a loan for college students L, a yearly loan repayment Γ and an income tax $T(y)$ that is constant over time, where these policy instruments satisfy*

- $T(y(a)) = y(a) - c(\theta_{HS}, a)$
- $\Gamma = c(\theta_{HS}, a) - c(\theta_{CO}, a)$
- $L = \beta^e(c_e + e)$.

[26] In contrast, in the case with income effects, education-independent marginal tax rates do not imply $y(\theta, a) = y(a)$. Concretely, as individuals with the same a but different θ will typically differ concerning their optimal consumption, they will choose different labor effort although they face the same labor wedge schedule. Imposing the same consumption and income level on all individuals with the same skill level a would overcome this problem, however, it would be a much stronger restriction on optimal policies. All these arguments do not necessarily imply that one cannot compute optimal history independent policies for the case with income effects. For this case, however, we would not be able to use a first-order approach but instead it would be necessary to check all possible incentive constraints. This would require us to significantly reduce the type space, severely limiting our ability to characterize non-linear schedules and make welfare comparisons across scenarios.

[27] Naturally, another implementation of this optimum would involve a single labor tax schedule with education-dependent lump-sums and education grants offered by the government.

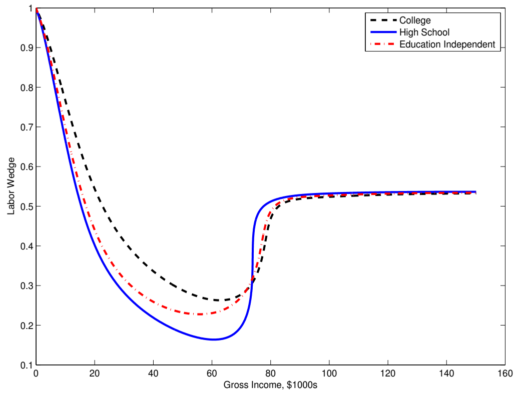


Figure 1.6: Optimal Education Independent Taxes

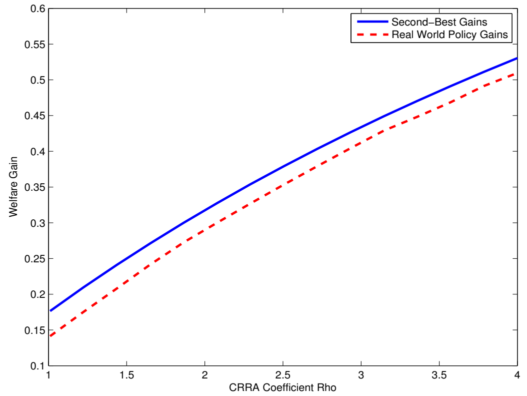


Figure 1.7: Welfare Gains

Figure 1.6 shows the optimal education independent labor income tax in this case; the optimal marginal tax rates lie between their education dependent counterparts from the second best optimum.

We next calculate the welfare gains from income-contingent repayment schemes for both cases: the unrestricted repayment schedule from Section 1.5.2 and the constrained one form Section 1.5.2.

In Figure 1.7 we present the consumption equivalent welfare gains as the CRRA parameter ρ varies from 1 to 4. First, one can see that the 'real world appeal'-repayment schedule is able to reap almost all the welfare gains from income-contingent loans. Second, the gains are increasing in risk-aversion which underscores the role of the loans as an insurance device. For a CRRA coefficient of two, the gains are about 0.32% in the unrestricted and about 0.25% in the restricted case. Thus, roughly 78% of the welfare gain from the second-best can be reaped with the restricted repayment. In case of an interest rate of 3%, 68% of the welfare gain can be reaped with simpler policies. For an interest rate of 5%, second-best optimal income-contingent repayment would actually never exceed the loan value.

Finally, the welfare gains are evenly distributed in the benchmark case ($\rho = 2$), implying that both the college and the high school type achieve a utility gain of 0.32% of consumptions equivalents. For lower values of ρ a larger share of the gain is reaped by the high-school graduates, for higher values of ρ the result is reversed.

1.5.3 Policies in Case of Differing Top Income Tails

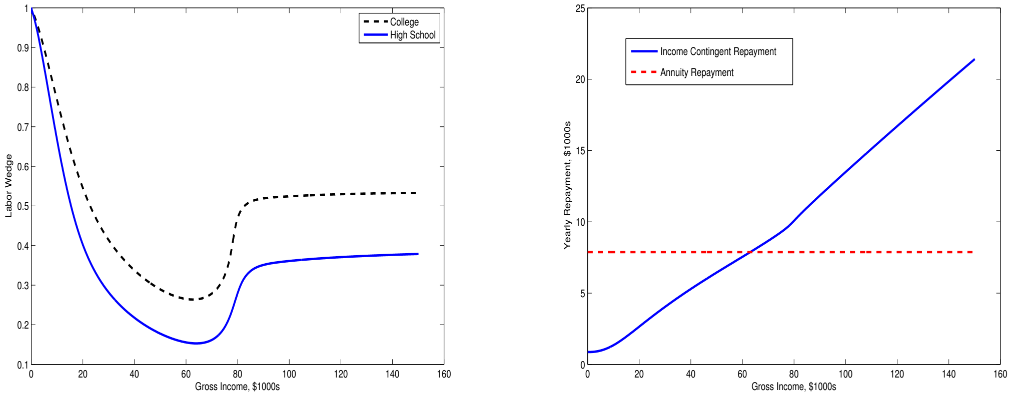
We now test if and how a different assumption on top incomes across income distributions changes the results. We focus on the case, where the college income distribution has a thicker tail than the high school income distribution. For college graduates, we choose a Pareto parameter of 1.28. For high-school graduates we choose a Pareto parameter of 3.[28] These values lie within the range of what has been typically found in empirical studies covering many countries and time periods (Atkinson, Piketty and Saez 2011). If we aggregate the two distributions to the aggregate income distribution, we find that the resulting tail for top incomes resembles a Pareto tail with a parameter not far away from 1.5.[29]

Second-Best Optimal Policies

Figures 1.8(a) and 1.8(b) display the corresponding schedules for labor wedges and the repayment schedule. The college labor wedge now lies above the high school labor wedge everywhere, leading to a strictly increasing repayment schedule. The implicit top tax rate for college graduates is higher than for high-school graduates, driven by the differences in the Pareto parameter. Interestingly, again a simple linear approximation of the repayment schedule with a linear slope of about 11% could almost perfectly implement the second-best optimum. Repayment of college graduates now exceeds the annuity loan value by a much more significant amount and for much bigger fraction of the population. We check again if a college type would prefer not to choose the income-contingent loan in this case and find that the loans indeed have to be compulsory. However, as we show next, one can again construct slightly different policies which respect a cap on repayment. These yield a large share of the welfare gain and do not require the loans to be compulsory.

[28]The top tails are not dependent on innate type θ but are just determined by the education level. In an earlier version of this paper (Findeisen and Sachs 2012), we also explore the case in which the tails are determined by innate type θ instead. The results are very similar.

[29]The sum of two Pareto distributions tends to behave like a Pareto distribution, where the heavier tail distribution seems to dominate (Ramsay 2006). This implies that, in the tails, the resulting aggregate distribution is very close to the college distribution.



(a) Labor Wedges (b) Repayment Schedule

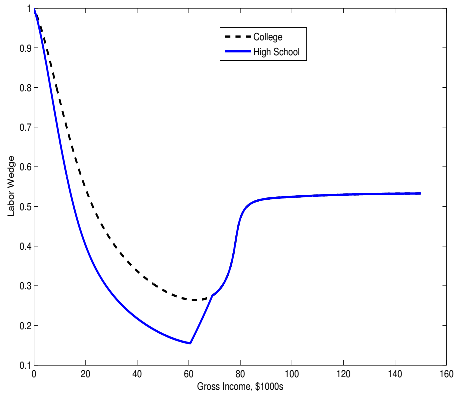
Figure 1.8: Utilitarian Optimum With Thick College Tails

Real World Polices: Cap on Repayment

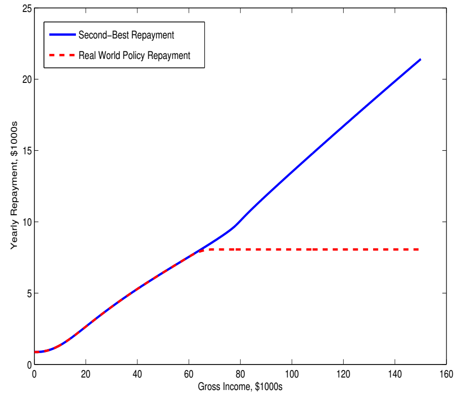
As in Section 1.5.2, we now adjust the second-best optimum towards policies that satisfy the same two mentioned real-world restrictions. The adjustment we make is slightly different this time. In Section 1.5.2, we lowered the labor wedges of the college types such that they equal the optimal ones for the high school types above all income levels, where the second-best repayment starts to exceed the loan value. Here, we do the opposite and increase the labor wedges of the high school types such that they are equal to the college labor wedges. The reason for this is that optimal history independent wedges (see Figure 1.10) are closer to the college wedges for high incomes, which is driven by the fatter college top income tail 'dominating' the top income tail for the high school types, see footnote 29. The new adjusted policies respect again incentive compatibility and budget feasibility. In order to avoid bunching because of a discrete upward jump in marginal tax rates, we smooth out the increase over an interval of roughly US-$5,000. The resulting labor wedges and repayment are illustrated in Figures 1.9(a) and 1.9(b).

The Welfare Gains From Income-Contingent Repayment

As in Section 1.5.2, we now calculate the welfare gains over students loans without income-contingent repayment. Due to the differing top income tails, the college and high school wedges are more distinct from each other (see Figure 1.10) than in the benchmark case. This yields to welfare gains (see Figure 1.11) that are slightly higher. They are 0.36% of lifetime consumption for a CRRA coefficient of 2. Again, the adjusted system respecting a cap can yield a large part of those gains: in fact, they lead to a gain



(a) Labor Wedges



(b) Repayment Schedule

Figure 1.9: Real World Optimum With Thick College Tails

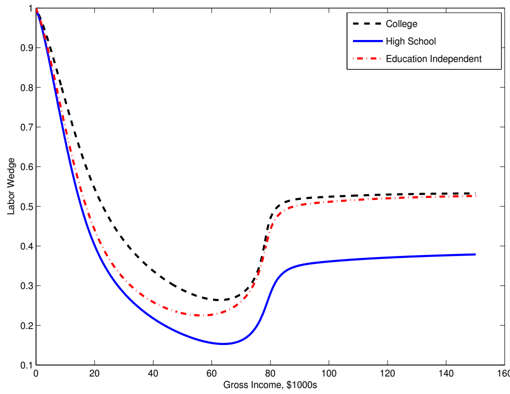


Figure 1.10: Optimal Education Independent Taxes

of 0.33%, which is almost 92% of the welfare gain. For an interest rate of 3% (5%) the latter value is 75% (95 %).

1.6 Conclusion

In this paper, we have studied the implications of endogenous education decisions before labor market entry on Pareto optimal tax policies in a dynamic environment with heterogeneous agents and uncertainty. An attractive way to decentralize Pareto optimal allocations is to have the government support students to finance consumption and tuition during education. During their working life students pay back these loans, contingent on income and loan size. Our paper therefore makes a second-best argument in favor of student loans with income-contingent repayment rates and, in addition, provides guidance for the optimal design of such repayment schedules.

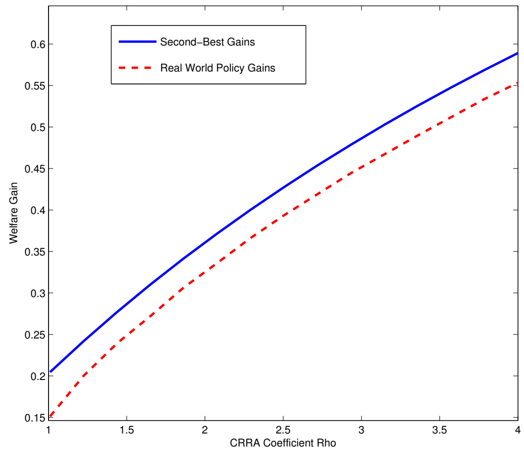


Figure 1.11: Skill Distributions

We have abstracted from several aspects that can be tackled in future work. First, we have abstracted from initial wealth heterogeneity. In an environment where individuals differ concerning the income and wealth of their parents, typically the question arises to what extent optimal education policies should depend on parents' income and wealth. Very related to this question, Gelber and Weinzierl (2012) have recently taken up the task of showing how the optimal history-independent tax system changes, when children's abilities depend on parents' financial resources. Second, due to our assumption that all labor market risk is realized directly after labor market entry, some aspects concerning the optimal timing of repayment were naturally disregarded. Relatedly, we did no consider human capital accumulation after labor market entry like on-the-job training.[30] Third, we assumed full commitment to all policies from the government side. Relaxing these assumptions might be a fruitful area for future research.

[30]In ongoing research, Stantcheva (2012) considers optimal taxation and human capital taxation in a life cycle economy, which encompasses on-the-job training.

1.7 Appendix – Chapter One

1.7.1 Proof of Lemma 1.3.1

Consider some admissible reporting strategy $r(\theta) = \theta'$.

$$\begin{aligned}\frac{\partial U(\theta,\theta')}{\partial r(\theta)} = u_c\left(c_e\left(\theta'\right)\right)\frac{\partial c_e(\theta')}{\partial r(\theta)} + \beta\int_{\underline{a}}^{\overline{a}}\frac{\partial v^w(\theta',a)}{\partial r(\theta)}g(a|e(\theta'),\theta)da \\ + \frac{\partial e(\theta')}{\partial r(\theta)}\int_{\underline{a}}^{\overline{a}}v^w(\theta',a)\frac{\partial g(a|e(\theta'),\theta)}{\partial z(\theta')}da\end{aligned}$$

and

$$\begin{aligned}0 = \frac{\partial U(\theta',\theta')}{\partial r(\theta)} = u_c\left(c_e\left(\theta'\right)\right)\frac{\partial c_e(\theta')}{\partial r(\theta)} + \beta\int_{\underline{a}}^{\overline{a}}\frac{\partial v^w(\theta',a)}{\partial r(\theta)}g(a|e(\theta'),\theta')da \\ + \frac{\partial e(\theta')}{\partial r(\theta)}\int_{\underline{a}}^{\overline{a}}v^w(\theta',a)\frac{\partial g(a|e(\theta'),\theta')}{\partial z(\theta')}da\end{aligned}$$

Subtracting from one another gives:

$$\begin{aligned}\frac{\partial U(\theta,\theta')}{\partial r(\theta)} = \beta\int_{\underline{a}}^{\overline{a}}\Bigg[\frac{\partial v^w(\theta',a)}{\partial r(\theta)}\left(g(a|e(\theta'),\theta) - g(a|e(\theta'),\theta')\right) \\ +\frac{\partial e(\theta')}{\partial r(\theta)}v^w(\theta',a)\left(\frac{\partial g(a|e(\theta'),\theta)}{\partial z(\theta')} - \frac{\partial g(a|e(\theta'),\theta')}{\partial z(\theta')}\right)\Bigg]da.\end{aligned}$$

We are now looking when this last expression always has the same sign as the difference $(\theta - \theta')$, which is clearly sufficient for global incentive compatibility. For $(\theta - \theta') > 0$, by Assumption 2, the first line is positive if $\frac{\partial v^w(\theta',a)}{\partial r(\theta)}$ or equivalently $\frac{\partial v^w(\theta,a)}{\partial \theta}$ in a truthful mechanism is increasing in a. This can be shown to be equivalent to $\frac{\partial y(\theta,a)}{\partial \theta} > 0$ using the envelope theorem. The second line is positive if $\frac{\partial e(\theta')}{\partial r(\theta)} > 0$ or equivalently $\frac{\partial e(\theta)}{\partial \theta} > 0$ in a truthful mechanism. The last condition (iii) is a routine exercise and a proof can be found, for example in Salanié (2003).

1.7.2 Optimal Labor and Education wedges

We start by stating the objective for the case of separable preferences of the form $u(c) - \Psi(l)$, where Ψ are the convex utility costs of labor. Further we assume that $u(\cdot) = u^e(\cdot)$. After integrating by parts and using the transversality conditions $\eta(\underline{\theta}) = \eta(\overline{\theta}) = 0$

as well as $\mu(\theta, \underline{a}) = \mu(\theta, \overline{a}) = 0 \ \ \forall \ \theta$, the Lagrangian for the social planner's problem reads as[31]

$$\begin{aligned}
\max_{c_e(\theta), v^w(\theta,a), e(\theta), y(\theta,a)} \mathcal{L} =& \beta^e \int_{\underline{\theta}}^{\overline{\theta}} u(c_e(\theta)) d\tilde{F}(\theta) \\
&+ \beta^w \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} v^w(\theta, a) dG(a|e(\theta), \theta) d\tilde{F}(\theta) \\
&+ \beta^w \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} y(\theta, a) dG(a|e(\theta), \theta) dF(\theta) \\
&- \beta^w \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} u^{-1} \left[v^w(\theta, a) + \Psi \left(y(\theta, a)/a \right) \right] dG(a|e(\theta), \theta) dF(\theta) \\
&- \beta^e \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \left(c_e(\theta) + e(\theta) \right) dF(\theta) \\
&- \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} \left(\mu'(\theta, a) v^w(\theta, a) + \mu(\theta, a) \Psi' \left(\frac{y(\theta, a)}{a} \right) \frac{y(\theta, a)}{a^2} \right) da d\theta \\
&- \int_{\underline{\theta}}^{\overline{\theta}} \eta'(\theta) \left[\beta^e u(c_e(\theta)) + \beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta, a) dG(a|e(\theta)) da \right] d\theta \\
&- \beta^w \int_{\underline{\theta}}^{\overline{\theta}} \eta(\theta) \int_{\underline{a}}^{\overline{a}} v^w(\theta, a) \frac{\partial g(a|e(\theta), \theta)}{\partial \theta} da d\theta
\end{aligned}$$

With first-order conditions:

$$u'(c_e(\theta))(\tilde{f}(\theta) - \eta'(\theta)) - \lambda_R f(\theta) = 0 \tag{1.8}$$

$$\begin{aligned}
&\left(\tilde{f}(\theta) - \eta'(\theta) \right) g(a|e(\theta), \theta) - \lambda_R \frac{1}{u'(c_w(\theta, a))} g(a|e(\theta), \theta) f(\theta) - \frac{\mu'(\theta, a)}{\beta^w} \\
&- \frac{\partial g(a|z(\theta), \theta)}{\partial \theta} \eta(\theta) = 0
\end{aligned} \tag{1.9}$$

[31] With more general preferences, the fourth line would be $\beta^w \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} \gamma \left(v^w(\theta, a), \frac{y(\theta,a)}{a} \right) dG(a|e(\theta), \theta) dF(\theta)$ with $\gamma(v, l)$ being the inverse function of u over c. The sixth line would be $-\beta^e \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \left(c_e(\theta) + e(\theta) \right) dF(\theta) - \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} \left(\mu'(\theta, a) v^w(\theta, a) + \mu(\theta, a) u_l \left(\gamma \left(v^w(\theta, a), \frac{y(\theta,a)}{a} \right), \frac{y(\theta,a)}{a} \right) \frac{y(\theta,a)}{a^2} \right) da d\theta$ instead.

$$\lambda_R g(a|e(\theta),\theta)f(\theta) - \frac{\mu(\theta,a)}{\beta^w}\left[\Psi''\left(\frac{y(\theta,a)}{a}\right)\frac{y(\theta,a)}{a^3} + \frac{1}{a^2}\Psi'\left(\frac{y(\theta,a)}{a}\right)\right]$$
$$- \lambda_R g(a|e(\theta),\theta)f(\theta)\frac{\Psi'\left(\frac{y(\theta,a)}{a}\right)}{au'(c_w(\theta,a))} = 0, \tag{1.10}$$

$$\tilde{f}(\theta)\beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)}da + \beta^w\lambda_R f(\theta)\int_{\underline{a}}^{\overline{a}} \frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)}\left(y(\theta,a) - c_w(\theta,a)\right)da$$
$$-\eta'(\theta)\beta^w\int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)}da - \beta^w\eta(\theta)\int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial^2 g(a|e(\theta),\theta)}{\partial e(\theta)\partial\theta}da - \beta^e\lambda_R f(\theta) = 0 \tag{1.11}$$

Combining equations (1.8) and (1.9) and integrating directly gives the inverse euler equation.

Proposition 1.3.2

Rewriting (1.10):

$$\lambda_R g(a|e(\theta),\theta)f(\theta)\left[1 - \frac{\Psi'\left(\frac{y(\theta,a)}{a}\right)}{au'(c_e(\theta,a))}\right]$$
$$-\frac{1}{\beta^w}\mu(\theta,a)\left[\Psi''\left(\frac{y(\theta,a)}{a}\right)\frac{y(\theta,a)}{a^3} + \frac{1}{a^2}\Psi'\left(\frac{y(\theta,a)}{a}\right)\right] = 0.$$

Dividing by $\frac{\Psi'}{au'}$ and $\lambda_R g(a|e,\theta)f(\theta)$ and using the definition of the labor wedge, i.e. $u'(1-\tau_y) = \Psi'\frac{1}{a}$ yields

$$\frac{\tau_y(\theta,a)}{1-\tau(\theta,a)} = \frac{1}{\beta^w}\frac{\mu(\theta,a)}{\lambda_R g(a|e(\theta),\theta)f(\theta)a}\left[\frac{\Psi''\frac{y}{a^2} + \Psi'\frac{1}{a}}{\frac{\Psi'}{au'}}\right],$$

which can be written as

$$\frac{\tau_y(\theta,a)}{1-\tau(\theta,a)} = \frac{1}{\beta^w}\cdot\frac{\mu(\theta,a)}{\lambda_R g(a|e(\theta),\theta)f(\theta)a}\frac{1+\varepsilon_u(\theta,a)}{\varepsilon_c(\theta,a)},$$

where $\frac{\Psi'\frac{1}{a}}{\Psi''\frac{y}{a^2}+\Psi'\frac{1}{a}} = \frac{1+\varepsilon_u(\theta,a)}{\varepsilon_c(\theta,a)}$ can be shown by simple algebra, see Saez (2001, p.227). In particular, with the isoelastic specification used in the computations $\frac{(y/a)^\sigma}{\sigma}$ one can verify that this term is equal to $\frac{1}{\sigma}$.

The multiplier $\mu(\theta,a)$ can be obtained using (1.9) and (1.8):

$$\frac{\mu(\theta,a)}{\beta^w} = \frac{\lambda_R f(\theta)}{u'(c_e(\theta))}G(a|e(\theta),\theta) - \lambda_R f(\theta)\int_{\underline{a}}^{a}\frac{1}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta)$$
$$-\frac{\partial G(a|e(\theta),\theta)}{\partial\theta}\eta(\theta^*),$$

yielding:

$$\frac{\tau_y(\theta,a)}{1-\tau(\theta,a)} = \frac{1+\varepsilon_u(\theta,a)}{\varepsilon_c(\theta,a)}\frac{u'(c_w(\theta,a))}{ag(a|e(\theta),\theta)}\left[\mathcal{A}(\theta,a)+\mathcal{B}(\theta,a)\right]$$

where

$$\mathcal{A}(\theta,a) = \frac{G(a|e(\theta),\theta)}{u'(c_e(\theta))} - \int_{\underline{a}}^{a}\frac{1}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta)$$

$$\mathcal{B}(\theta,a) = -\frac{1}{f(\theta)\lambda_R}\frac{\partial G(a|e(\theta),\theta)}{\partial\theta}\eta(\theta).$$

Using the inverse Euler equation, the term $\mathcal{A}(\theta,a)$ can be written as in the proposition. From (1.8), $\eta(\theta)$ is given by:

$$\eta(\theta) = \tilde{F}(\theta) - \lambda_R\int_{\underline{\theta}}^{\theta}\frac{1}{u'(c_e(\theta))}f(\theta)d\theta.$$

The direct benefit of raising utils for agents with skill lower than θ is $\tilde{F}(\theta)$. The monetary cost is $\int_{\underline{\theta}}^{\theta}\frac{1}{u'(c_e(\theta))}f(\theta)d\theta$, transformed into welfare units by λ_R.

Relation to the formula of Saez (2001)

The insurance part of the labor wedge can be expressed as in Saez (2001), for our case with separable preferences. This relation applies if agents do not differ ex-ante. By some abuse of notation, then $\mathcal{B}(\theta,a)=0$ and for $\mathcal{A}(\theta,a)$, using the inverse Euler equation, we obtain

$$\begin{aligned}
\mathcal{A}(\theta,a) &= \int_{\underline{a}}^{\overline{a}}\frac{G(a|e(\theta),\theta)}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta) - \int_{\underline{a}}^{a}\frac{1}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta)\\
&= \int_{\underline{a}}^{\overline{a}}\frac{G(a|e(\theta),\theta)}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta) - \int_{\underline{a}}^{\overline{a}}\frac{1}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta)\\
&+ \int_{a}^{\overline{a}}\frac{1}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta)\\
&= \int_{a}^{\overline{a}}\frac{1}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta) - \int_{\underline{a}}^{\overline{a}}\frac{1-G(a|e(\theta),\theta)}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta)
\end{aligned}$$

where the second equality follows from the transversality condition. This term can be expressed as in Saez (2001) as shown by Mankiw, Weinzierl and Yagan (2009) in their online appendix.

General Preferences: Carrying out the analogous steps with a general utility function $u(c, l)$ we get:

$$\frac{\tau_y(\theta,a)}{1-\tau(\theta,a)} = \frac{1+\varepsilon_u(\theta,a)}{\varepsilon_c(\theta,a)} \frac{u_c(\theta,a)}{ag(a|e(\theta),\theta)} \int_a^{\overline{a}} \exp\left(-\int_a^x \frac{u_{c,l}(\theta,s)}{u_c(\theta,s)} \frac{l(\theta,a)}{a} ds\right) \times [\mathcal{A}(\theta,x)+\mathcal{B}(\theta,x)]\, dx,$$

where $\mathcal{A}(\theta,x) = g(x|e(\theta),\theta)\left(\frac{1}{u_c(c_w(\theta,x))} - \frac{1}{u_c(c_e(\theta))}\right)$ and $\mathcal{B}(\theta,x) = \frac{\eta(\theta)}{\lambda f(\theta)} \frac{\partial g(x|e(\theta),\theta)}{\partial \theta}$.

Proposition 1.3.3

Plugging 1.8 into 1.11 gives:

$$\frac{\lambda_R f(\theta)}{u'(c_e(\theta))}\beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)} da + \beta^w \lambda_R f(\theta) \int_{\underline{a}}^{\overline{a}} \frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)} \left(y(\theta,a) - c_w(\theta,a)\right) da$$
$$- \beta^w \eta(\theta) \int_{\underline{a}}^{\overline{a}} v^w(\theta,a) \frac{\partial^2 g(a|e(\theta),\theta)}{\partial e(\theta)\partial\theta} da - \beta^e \lambda_R f(\theta) = 0$$

Proposition 1.3.3 directly follows. Note that the relevant first-order conditions are identical for general utility function, so that the formula for the optimal wedge in the proposition also applies.

1.7.3 Implementation Appendix

Proof of Proposition 1.4.1

Starting from a direct mechanism we show in four steps that optimal allocations can indeed be implemented with the policy instruments as defined in Proposition 1.4.1. The idea to work with a history independent savings tax builds upon the work of Werning (2011).

Step 1: Introduce savings

Imagine the constrained efficient allocation is implemented by a direct mechanism. From that point on, assume that individuals could freely save s at rate R. Let r_1 denote the report about θ. With savings tax $T^s(s, r_1)$, the budget constraints read as

$$\tilde{c}_e(r_1) + s = c_e(r_1) \tag{1.12}$$

$$\tilde{c}_w(r_1, r_2) = c_w(r_1, r_2) + Rs - T^s(s, r_1) \tag{1.13}$$

Define the optimal report r_2 about a, for a given report r_1 about θ, a given savings tax schedule $T^s(s, r_1)$ and a given level of savings s:

$$r_2^*(a, r^1, s, T^s) = \arg\max_{r_2} u \left[c_w(r_1, r_2) + Rs - T^s(s, r_1) - \psi\left(\frac{y(r_1, r_2)}{a}\right)\right]$$

Then the optimal report in period one, for a given level of savings and a given savings tax schedule $T^s(s, r_1)$, is defined by

$$\begin{aligned} r_1^*(\theta, s, T^s(r^1, s)) &= \arg\max_{r_1} u(c_e(r_1) - s) \\ &+ \beta \int_{\underline{a}}^{\overline{a}} u \left[c_w(r_1, r_2^*) + Rs - T^s(s, r_1) - \psi\left(\frac{y(r_1, r_2^*)}{a}\right)\right] dG(a|e(r^1), \theta) \end{aligned} \tag{1.14}$$

Then define a hypothetical tax schedule $T^*(r_1, s, \theta)$ for each θ implicitely by[32]

$$V(\theta) = V(\theta, s, r_1^*, T^*(r_1, s, \theta)) \; \forall \; s.$$

This hypothetical tax schedule would make individuals of type θ indifferent between truth telling and the optimal joint deviation for any s. It is hypothetical since it does not only depend on the report r_1, which is observable but also on the unobservable type θ. However, we know that for each θ such a tax schedule exists. Therefore taking the upper envelope over these functions yields a savings tax function $\hat{T}(s, r_1)$ that also implements zero savings and is feasible since it does not condition on θ:

$$\hat{T}(s, r_1) = \sup_{\theta} T^*(r_1, s, \theta). \tag{1.15}$$

Lemma 1.7.1. *A constrained efficient allocation can be implemented by a direct mechanism extended by a savings choice and history-dependent savings tax schedules $\hat{T}^s(s, r_1)$.*

Step 2: Make the savings tax history-independent

A simple way to make the savings tax history-independent is to take the upper envelope of all functions $T^s(s, r_1)$, i.e.

$$T^s(s) = \sup_{r_1} \hat{T}^s(s, r_1). \tag{1.16}$$

Lemma 1.7.2. *A constrained efficient allocation can be implemented by a direct mechanism extended by a savings choice and a history-independent savings tax schedule $T^s(s)$.*

Note that this savings tax function $T^s(s)$ is not differentiable and implies zero savings. As Werning (2011) shows one can, using Ricardian equivalence arguments, also

[32] Recall that $V(\theta)$ is the value function of a truth teller of type θ.

construct a history-independent savings tax function that is differentiable and yields non-zero savings choices.

Step 3: Allow for labor-leisure decisions

To get closer to a decentralized implementation now assume the following extended direct mechanism.

1. Individuals report $r(\theta)$
2. They get assigned 'income to consume' $c_e(\theta)$
3. They face the savings tax schedule $T^s(s)$ and save $s(\theta) = 0$
4. In period two, instead of directly revealing their type, individuals of type θ face an income tax schedule that is defined by

$$T(y(\theta, a), e(\theta)) = y(\theta, a) - c_w(\theta, a) \; \forall \; a.$$

By the same arguments as in the standard Mirrlees model it follows that this extended direct mechanism can also implement the constrained efficient allocations. We can summarize this in the following lemma.

Lemma 1.7.3. *A constrained efficient allocation can be implemented by a direct mechanism in the first period extended by a savings choice and a history-independent differentiable savings tax schedule $\tilde{T}^s(s)$ and a history-dependent labor income tax schedule $T(y, e)$ in period two.*

Step 4: Complete Decentralization – allow for educational investment

1. Individuals buy (or tell the government that they want to buy) $e(\theta)$ units of education
2. They get assigned a student loan $\mathcal{G}(e(\theta)) = c_e(\theta) + e(\theta)$ (and are obliged to actually buy $e(\theta)$ units of education)
3. They face the savings tax schedule $T^s(s)$ and save $s(\theta) = 0$
4. in period two, instead of directly revealing their type, individuals of type θ face an income tax schedule that is defined by

$$T(y(\theta, a), e(\theta)) = y(\theta, a) - c_w(\theta, a) \; \forall \; a$$

Since the mechanism in step 4 is just a reformulation of the mechanism in step 3 this directly leads us to Proposition 1.4.1.

Discussion of Implementations with Non-Monotone Education

If education is not strictly monotone, it is not enough to condition tax and grant schedules on education for education levels which are assigned to more than one type. In this case for those respective education levels, the planner can augment the system of education grants, such that there is no more a unique grant per education level but a set of grants, which contains the respective correct grant. More formally, if the education level e^* is the optimal education level for all individuals within a certain set $\Theta(e^*)$, then the set of grants assigned to education level e^* must contain $\mathcal{G}(e^*, \theta^*) = c_e(\theta^*) + e^*$, for each $\theta^* \in \Theta(e^*)$. Every education dependent tax function associated with a level e^*, can then in addition be conditioned on consumption during education $c_e(\theta^*)$. Analogously, the planner can offer multiple loans sizes per education level, and repayment schedules which condition on the loan size, income and early consumption.

Chapter 2

The Rise of the East and the Far East: German Labor Markets and Trade Integration[1]

This chapter is joint work with Wolfgang Dauth and Jens Suedekum. It is a revised version of IZA Working Paper No. 6685.

It is submitted for publication to the *Journal of the European Economic Association*.

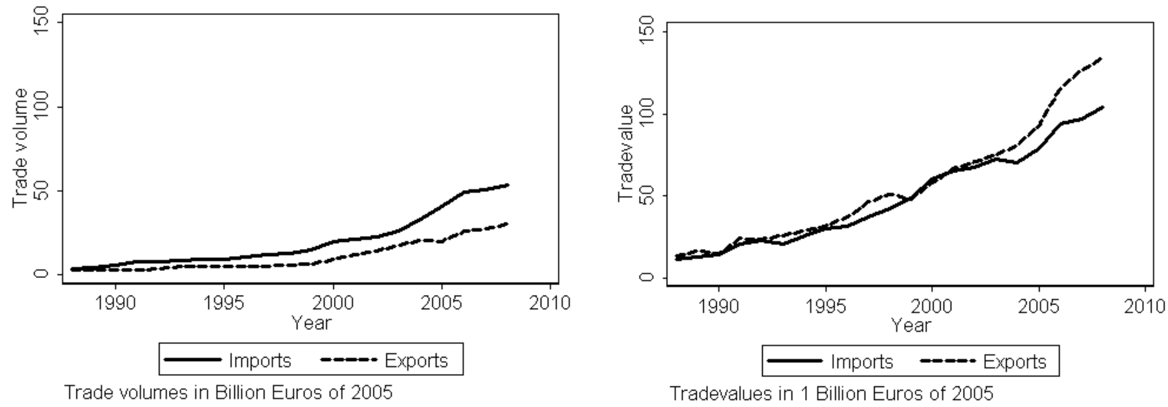
2.1 Introduction

One of the central forces of globalization in the last decades is certainly the rise of Eastern Asian countries, especially China, in the world economy. The substantial rise of trade with China, and its perceived competitiveness, have led to major concerns in Western market economies about possible adverse effects for domestic labor markets. This "fear" is particularly high on the agenda in the United States, and numerous studies have addressed the impacts of this trade integration on the US economy.[2]

[1]Email contacts: wolfgang.dauth@iab.de, sebastian.findeisen@uzh.ch, jens.suedekum@uni-due.de. We thank David Autor, Marius Brülhart, Bernd Fitzenberger, William Kerr, Fabrizio Zilibotti and participants of various seminars for helpful comments. The usual disclaimer applies.

[2]See, among others, Feenstra and Hanson (1999); Harrigan (2000); Feenstra and Wei (2010); Harrison, McLaren, and McMillan (2010); Ebenstein, Harrison, McMillan, and Phillips (2011).

From the perspective of Germany, which consistently ranks among the most open economies in the world and for a long time held the unofficial title of the export world champion, China's rise also had a major impact. Starting from almost zero trade in the late 1980s, the German import volume from China has risen dramatically to more than 50 billion Euros in 2008 (see Figure 2.1).



(a) China (b) Eastern Europe

Figure 2.1: German trade volumes with China and Eastern Europe, 1988-2008.

This corresponds to a growth rate of 1608 percent, which is far higher than for any other trading partner (see Table 2.1). However, although Germany runs a trade deficit

Table 2.1: Changes in German trade volumes, 1988-2008 (in Billion Euros of 2005)

	China		Eastern Europe	
Period	Imports	Exports	Imports	Exports
1988	3.1	3.0	11.0	13.3
1998	12.9	5.6	42.0	51.0
2008	53.1	30.1	103.8	134.0
Growth	**1628.3%**	**893.2%**	**843.9 %**	**905.3%**

	Other Asian dev. countries		Rest of the World	
Period	Imports	Exports	Imports	Exports
1988	5.0	5.1	289.4	402.1
1998	12.5	7.5	357.7	506.9
2008	20.0	16.3	490.2	842.7
Growth	296.5 %	219.0 %	69.4 %	109.6 %

Source: Own calculations based on United Nations Commodity Trade Statistics.

vis-a-vis China despite an overall trade surplus, the magnitude of this deficit is much smaller than in the US case. This is because German exports to China have also risen by about 900 percent, from almost zero in 1988 to some 30 billion Euros in 2008. The "rise of China" therefore led to two major impacts for the German economy: Increased import competition particularly in such sectors as textiles, toys, or lower-tier office and computer equipment, but at the same time a substantial rise in market opportunities for German export sectors, most notably automobiles, specialized manufacturing, and the electronic and medical industries.

In addition to the "rise of China", Germany was affected by another major facet of globalization that at least economically had a much milder impact in North America, namely the fall of the Iron Curtain with the subsequent transformation of the former socialist countries into market economies. Overall, the rise of German exports to Eastern Europe even outpaced export growth to China. Import growth from Eastern Europe also has been substantial, exceeding 800 percent during the period 1988-2008.[3] For the German economy, import competition and export market opportunities therefore increased not only from the Far East, but also from the East closer by.

In this paper, we analyze the impacts of these major trade liberalizations from the perspective of small-scale German regions. There is substantial variation in sectoral employment patterns at the regional level, also within the manufacturing sector where commodity trade occurs. Given these initial specializations, regions are differently exposed to import competition and export opportunities arising from Eastern European and Asian countries. Regions that are strongly specialized in export-oriented industries, say "automobile regions", may benefit from the rise of new markets, while regions specialized in import industries, say "textile regions", may see their labor markets put under strain by the rising exposure to foreign competition. In our aggregate analysis, we relate changes in key local labor market variables to measures of import and export exposure that reflect the local industry mix. Afterwards, we adopt a complementary, more disaggregate approach at the level of individual workers, analyzing how trade exposure affects employment stability within regions, local industries, and plants.

In the literature, there are several approaches to identifying the impacts of trade shocks. One approach uses industries at the national level as the unit of observation and analyzes the general equilibrium impacts of trade, taking into account that inter-sectoral labor mobility may also involve a loss of specific human capital (Feenstra and Hanson (1999), Harrigan (2000), Robertson (2004), Poletaev and Robinson (2008), Blum

[3] To obtain a geographically stable region, we consider Eastern Europe to comprise the countries Bulgaria, Czech Republic, Hungary, Poland, Romania, Slovakia, Slovenia, and the former USSR or its succession states Russian Federation, Belarus, Estonia, Latvia, Lithuania, Moldova, Ukraine, Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan. The increase in trade volumes between the US and these countries is negligible, at least in comparison to the German numbers. The sectoral structure of German trade with Eastern Europe differs from trade with China – see Tables 2.2 and 2.3 in the Appendix. Although the export sectors are mostly the same, there is more intra-industry and vertical trade as the top imported items are automobile parts and electric apparatus.

(2008)). This literature is based on the view that labor markets adjust instantaneously or very rapidly to a new equilibrium, even after major perturbations. Another prominent approach looks at finer levels of disaggregation and is based on the presumption that the adjustment to major trade shocks is sluggish and may require more time. In that case, the differential impacts on firms, occupations or regions may be informative about the short- to medium-run effects of trade liberalization. Within that string of literature, Bernard, Jensen, and Schott (2006), Verhoogen (2008), Amiti and Davis (2012), and Bloom, Draca, and Van Reenen (2011) have analyzed trade shocks at the level of plants and firms, whereas Artuc, Chaudhuri, and McLaren (2010), McLaren and Hakobyan (2010), and Ebenstein, Harrison, McMillan, and Phillips (2011) use the industry and occupation level.

Very recently, a literature has started that identifies the impact of trade shocks at the *regional level*, thereby addressing the intra-national impacts of inter-national trade integration. This paper is most closely related to this string of literature, see Chiquiar (2008), Kovak (2011), Topalova (2010), Bruellhart, Carrere, and Trionfetti (2012), and in particular, Autor, Dorn, and Hanson (2012) (henceforth labeled as *ADH*). Interestingly, although our empirical approach is similar to *ADH*, we obtain results for Germany that differ substantially from their findings for the US, and that have very different implications about the overall impact of globalization for the domestic economy. In particular, our results strongly point towards net employment gains from increased trade with the East, drawing a very different picture for Germany than for the US. These differences arise from different structures within the manufacturing sector, at the onset and during the process of economic reform in China and Eastern Europe.[4]

ADH separate the US into 722 labor market regions and analyze the differential performance of these regions depending on their exposure to import competition from China. To account for unobserved shocks that simultaneously affect imports and regional performance, they use imports of other high-income countries to construct an instrument for US regional import exposure. Their main finding is that regions strongly exposed to Chinese import competition have experienced severe negative impacts on their labor markets, such as rising unemployment or lower labor force participation.[5] Importantly, this negative impact even seems to prevail when taking into account that the rise of China also implies new market opportunities for US producers. That is, the

[4]Given the enormous differences in industry structure between the US and Germany, the experience of those countries seem to lie on opposing ends of the spectrum. Our conclusions for Germany may be representative for other developed economies to the extent that they also specialize in modern, export-oriented manufacturing goods.

[5]ADH also find that Chinese trade shocks induced only small cross-regional population shifts. This low labor mobility, in turn, supports the view that regions can be treated as "sub-economies" across which adjustment to shocks works far from instantaneously. Since regional labor mobility in Germany is traditionally much lower than in the US Molloy, Smith, and Wozniak (2011), Bertola (2000), that approach indeed seems especially well applicable in our context.

impact of local export exposure on labor market performance across US regions appears to be weak, and did not compensate the adverse impacts of import penetration.

In our empirical analysis, we pay particular attention to the overall exposure of German regions to trade with "the East" – that is, China and Eastern Europe – both from the import and from the export side. The rise of China, facilitated by substantial productivity gains and the Chinese WTO accession, and for that matter also the rise of Eastern Europe that was due to similar causes, not only imply an exogenous increase in import exposure from the point of view of a single German region; they also imply an increase in new export opportunities that regions specialized in the "right" type of industries can take advantage of. Our results suggest that this latter aspect is in fact crucial for understanding how German local labor markets were affected by, and adjusted to trade exposure in the past two decades. Consistent with *ADH*, we also find a negative *causal* effect of import exposure on manufacturing employment in German regions. That is, regions specialized in import competing sectors saw a decline in manufacturing employment attributable to the impact of trade. Yet, this negative impact is on average offset by a positive *causal* effect of export exposure, as the respective export oriented regions built up manufacturing employment as a result of the new trade opportunities. In addition, we find that trade integration with Eastern Europe had a much bigger impact on Germany than the rise of China.

In the aggregate, we therefore find that the "rise of the East" has created jobs in the German economy. A back-of-the-envelope calculation quantifies this effect to range around 493,000 full-time equivalent jobs in the period 1988–2008 that would not have existed without the trade integration. This aggregate implication is very different from ADH's conclusions for the US, and we discuss some possible explanations (such as the overall trade balance) for these differences below.

We also move beyond the manufacturing sector, and investigate how local labor markets responded more broadly to the increase in trade exposure. Here we shed light on questions such as: what happens to the workers displaced by trade exposure, or to what extent do the trade effects spill over to other (non-manufacturing) sectors in the economy? We find that regions specialized in export industries saw significant total employment gains and reductions in unemployment. Those gains clearly occur within the manufacturing sector, which is retained in Germany as a result of the deepening of trade, but employment in complementary business related services (such as accounting or consulting) also gained notably. Import-competing regions, on the other hand, were affected adversely also beyond the manufacturing sector.

Finally, our analysis at the individual level allows for an even more detailed look on the causal effects of trade. Here, we use cumulative spell information from administrative social security data. We find that a higher export exposure of the own job raises the probability of staying employed in the same plant or local industry. Analogously, higher import exposure raises the probability that a job is terminated. Overall, however, we find that trade has led to a higher stability of employment relationships.

The rest of this paper is organized as follows. Section 2.2 describes the empirical approach. Section 2.3 is devoted to the analysis of manufacturing employment at the regional level, while Section 2.4 looks at further regional labor market outcomes. Section 2.5 presents the worker level analysis, and Section 2.6 concludes.

2.2 Estimation Strategy

2.2.1 Trade Exposure Across Local Labor Markets

Our empirical strategy is linked to the approach by *ADH* which exploits the variation in initial industry specialization across local labor markets at the onset of the economic rise of a trading partner, in our context Eastern Europe and China.

We first consider the import exposure of a German region i from "the East". Using *ADH*'s approach, which is based on a monopolistic competition model of international trade with cross-country productivity differences, this import exposure can be written as follows:

$$\Delta(ImE)_{it}^{EAST} = \sum_j \frac{E_{ijt}}{E_{jt}} \frac{\Delta Im_{jt}^{EAST}}{E_{it}}, \tag{2.1}$$

where ΔIm_{jt}^{EAST} is the total change in imports from the East to Germany (in constant Euros of 2005) that was observed in industry j between time periods t and $t+1$.[6] E_{ijt}/E_{jt} represent region i's share of national industry employment in j, and E_{it} is total manufacturing employment in period t and region i. This measure thus captures the *potential* increase in import exposure of a region i given its initial sectoral employment structure, as it apportions the *national* change in imports to the single German regions according to the regions' shares in national industry employment.

Figure 2.1 in the Appendix illustrates this import exposure for the period 1998 to 2008, both with respect to China and Eastern Europe. As can be seen from the maps, there is strong variation in these exposure measures, reflecting substantial differences in sectoral structures across regions. It stands out that the industrial structure of Eastern Germany in 1998 was such that there was little potential import competition, neither from China nor from Eastern Europe. The West was, by and large, exposed more strongly although there is substantial regional variation. Notice also that the correlation between Chinese and Eastern European import exposure is only about 0.3. That is, many German regions were exposed quite strongly to imports from one area, but not from the other. The

[6]In the benchmark specification below we consider that China and Eastern Europe together form "the East", so that ΔIm_{jt}^{EAST} refers to the joint increase of German imports from both areas. In further specifications, we consider import exposure from China and Eastern Europe separately.

average increase in exposure to Chinese imports over that time period was €1,903, while for Eastern Europe it was €1,848.

To capture regional export exposure, we derive an analogous measure:

$$\Delta(ExE)_{it}^{EAST} = \sum_j \frac{E_{ijt}}{E_{jt}} \frac{\Delta Ex_{jt}^{EAST}}{E_{it}}, \tag{2.2}$$

which captures the potential of regions, given their initial sectoral employment patterns, to benefit from rising demand from the "East" for German manufacturing products. Figure 2.2 in the Appendix illustrates the increase in potential export exposure of German regions, both with respect to China and Eastern Europe. The mean export exposure for China was €1,037, while for Eastern Europe that number reached €3,714. The map again shows that Eastern Germany is relatively little affected, while there is substantial regional variation in the West, yet with a clearly visible concentration in the southern and south-western part where the automobile and machinery sectors are highly concentrated.

2.2.2 Instrumental Variable Strategy

In the empirical analysis we aim to identify the causal effect of the rise of the East on the economic performance of German regions. More specifically, we regress the change of regional manufacturing employment, and other variables, between t and $t+1$ on the change of regional import and export exposure over the same time period.

The main challenge for this exercise is the endogeneity of trade exposure, in particular the presence of unobserved supply and demand shocks that simultaneously affect import/export exposure and regional economic performance. To address these concerns, we employ an instrumental variable (IV) strategy that is close in spirit to the approach by *ADH*. To instrument German regional import exposure from the East, we construct the following variable for every German region i:

$$\Delta(ImE_{Inst})_{it}^{EAST} = \sum_j \frac{E_{ijt-1}}{E_{jt-1}} \frac{\Delta Im_{jt}^{EAST-other}}{E_{it-1}}. \tag{2.3}$$

Here, $\Delta Im_{jt}^{EAST-other}$ are changes in trade flows of industry j's goods from the East (China and Eastern Europe) to other countries (see below). Similarly, for regional export exposure we construct the following instrumental variable that uses changes in exports of other countries to China and Eastern Europe:

$$\Delta(ExE_{Inst})_{it}^{EAST} = \sum_j \frac{E_{ijt-1}}{E_{jt-1}} \frac{\Delta Ex_{jt}^{EAST-other}}{E_{it-1}}. \tag{2.4}$$

The identification strategy (2.3) is based on the idea that the rise of Eastern Europe/China in the world economy induces a supply shock and rising import penetration for *all* trading partners, not just for Germany. Constructing a regional measure of import exposure by using those import flows of other countries therefore identifies the exogenous component of rising competitiveness in the East, and purges the effects of possible other shocks that simultaneously affect German imports and regional performance variables.[7] The logic of the instrumental variable (2.4) is similar. As the East rises in the world economy, it becomes a more attractive export destination for *all* countries, not just for Germany. Using (2.4) as an instrument for (2.2) thus purges the impacts of other unobservable shocks, and thus identifies the causal impact of the rise of export opportunities in the East on German local labor markets.

The quality of the instruments hinges, in particular, on three important conditions. First, they must have explanatory power in order to avoid a weak instrument problem. Second, the supply and demand shocks in those other countries should not be too strongly correlated with those of Germany, since otherwise the instruments do not purge the internal shocks so that the estimated coefficients are still biased. Third, in order for the exclusion restriction not to be violated, there should not be an independent effect of the trade flows of those other countries with China and Eastern Europe on the German regions, other than through the exogenous rise of the East.

To take those conditions into account, it is important to consider which countries are included in the "instrument group" whose trade flows are used to construct (2.3) and (2.4). We adopt the following approach: We focus on developed countries with a similar income level as Germany, but we exclude all direct neighbors as well as all members of the European Monetary Union. This is for two reasons. First, supply and demand shocks in such countries (e.g., France or Austria) are likely to be too similar to those in Germany, hampering the identification. Second, since those countries are highly integrated with Germany in an economic union where exchange rate alignments are impossible, it is likely that shocks which change trade flows between those countries and China/Eastern Europe also directly affect regional performance in Germany. We also do not consider the United States in the instrument group, because of its high significance in the world economy that is likely to violate the exclusion restriction. Our final "instrument group" consists of Australia, Canada, Japan, Norway, New Zealand, Sweden, Singapore, and the United Kingdom. Below we conduct several robustness checks where we change the countries that are included in the instrument group.

[7]Notice that the import values of the other countries are distributed across the German regions according to *lagged* sectoral employment shares from period $t-1$. This is done in order to tackle potential issues of measurement error or reverse causality, if employment reacted to anticipated trade. In practice using lagged or contemporaneous employment to construct the instrument turns out to have no significant impact on the results.

2.3 Trade exposure and Manufacturing Employment

2.3.1 Data

For the analysis at the regional level, we combine two main data sources. The German labor market data at the regional and local industry level come from the IAB-Establishment History Panel (BHP, see Spengler (2008)) which includes the universe of all German establishments with at least one employee subject to social security. This data set consists of an annual panel with approximately 2.7 million yearly observations on establishments aggregated from mandatory notifications to social security in the years from 1975 to 2008. Due to the administrative origin, the data are restricted to information relevant for social security (structure of workforce with regard to age, sex, nationality, qualification, occupation, wage) but at the same time are highly reliable and available on a highly disaggregated level.

Detailed data for regional sectoral employment is available from 1978 onwards. Since much of the rise of China and Eastern Europe occurred after 1990, we use 1988 as our starting point and thus observe data for two time periods (1988 to 1998 and 1998 to 2008) for each region. This timing also allows us to use employment lagged by ten years in the construction of our instruments as discussed above. Eastern German regions are only included for the second decade 1998 to 2008, because sectoral employment data for these regions only became available in the mid-1990s.

Information on international trade is taken from the United Nations Commodity Trade Statistics Database (Comtrade). This data contains annual international trade statistics of over 170 reporter countries detailed by commodities and partner countries. Trade flows are converted into Euros of 2005 using exchange rates supplied by the German Federal Bank. We merge these two data sources by harmonizing industry and product classifications. The correspondence between 1031 SITC rev. 2/3 product codes and the employment data (101 NACE 3-digit equivalent industry codes) is provided by the UN Statistics Division and allows unambiguously matching 92 percent of all commodities to industries. Trade values of ambiguous cases are partitioned into industries according to national employment shares in 1978.

2.3.2 Baseline Specification: Manufacturing Employment Growth

We estimate the effect of trade exposure on local labor markets by running specifications of the form:

$$\Delta Y_{it} = \gamma_t + \beta_1 \Delta (ImE)_{it}^{EAST} + \beta_2 \Delta (ExE)_{it}^{EAST} + X_{it}' \beta_3 + e_{it}. \qquad (2.5)$$

That is, we relate changes in the regional outcome variable Y_{it} between time periods t and $t+1$ to changes in (potential) regional import and export exposure from the East (i.e., Eastern Europe and China) during the same time period, while controlling for start-of-period regional control variables X'_{it}. In the baseline specification of this section, the dependent variable is the decennial change in manufacturing employment as a share of the working age population in region i, $Y_{it} = E_{it}^{M/WP}$. In the next section we consider further outcome variables.[8]

In the most parsimonious specification the vector X'_{it} includes dummies for the 16 German federal states and a time dummy γ_t to capture decade specific trends. Furthermore, we control for the overall regional employment shares of tradeable goods industries since our approach exploits the detailed regional variations of employment structures *within* the manufacturing sector. In more comprehensive specifications, we then add further controls for the initial composition of the local labor force, namely the start-of-period share of high-skilled workers, foreigners and women. Furthermore, motivated by the literature on job off-shoring, e.g. Antras, Garicano, and Rossi-Hansberg (2006) and Grossman and Rossi-Hansberg (2008). We include the percentage of routine intensive occupations (represented by basic activities in the taxonomy of Blossfeld (1987)). Table 2.1 in the appendix reports some descriptive statistics for the main variables.

Main results The first three columns of Table (2.2) show OLS specifications where we do not instrument for import and export exposure. Column 1 includes only the parsimonious set of controls. As can be seen, export exposure is estimated to have a positive and significant relationship with manufacturing employment growth, whereas the relationship with import competition is not statistically different from zero. We also find a trend of mean reversion of manufacturing employment, since growth is negatively related to the initial employment share of tradeable goods industries. In column 2 we add the further regional control variables, and we find that this leaves the results for the central variables (import and export exposure) unaffected. The coefficients for those other controls have the expected sign: A higher share of high-skilled, foreign and female workers in the local labor force is negatively related to manufacturing employment growth, since those groups are more prevalent in service industries. For the share of routine occupations we find no clear relationship. Finally, in column 3 we use interacted federal state $\times$ time period dummies instead of separate state/time dummies. This specification is the most demanding one, as it is only identified by within state-time variation. As can be seen, the coefficients for trade exposure as well as for the other control variables remain stable.

[8]To account for spatial and serial correlation, we cluster the standard errors at the level of 50 high-order labor market areas as defined in Kropp and Schwengler (2011) in all specifications.

The OLS coefficients reported in the first three columns are confounded with unobservable supply and demand shocks that can simultaneously affect employment and trade flows in Germany. To address this bias, we now turn to the IV strategy described before. When using the instrumental variables (2.3) and (2.4) for (2.1) and (2.2), we find that the impact of import exposure is now both statistically and economically highly significant. The results indicate that the sources of bias for the OLS estimates of import exposure seem to be quantitatively important and responsible for driving the OLS estimates towards zero.[9] The coefficient for export exposure, on the other hand, remains in the same ballpark as before. Table 2.2 also reports the Kleibergen-Paap Wald rk F statistic to diagnose a potential weak instrument problem.[10] With values in the order of 20, the results suggest that we face no such weak instrument bias – the values are well above the critical values compiled by Stock and Yogo (2002).[11] To further examine the explanatory power of our instruments, table 2.4 in the appendix reports details on the first stage regressions. As can be seen, both excluded instruments explain import exposure, the t-values of the imports-instrument being larger than the ones of the exports-instrument. Export exposure is exclusively explained by the exports-instrument. This further corroborates the credibility of our instruments relying on the assumption that the increasing trade exposure of German regions due to the rise of the East can be explained by Eastern trade with other countries. Since we use a just-identified model, it is not straightforward to examine the exogeneity of the excluded instruments. Yet, in section 2.3.3, we specify an overidentified model and include two instruments for each country. This allows us to compute the Hansen's J test which does not indicate correlation of the error term and the instruments.

Eastern Europe versus China The results so far refer to the joint impact of trade exposure with China and Eastern Europe. In Table (2.3) we consider the impact of trade exposure separately for Eastern Europe and China. We henceforth only report the IV estimates for the same three specifications as in columns 4 to 6 of Table (2.2), and for brevity we focus on the results for the main variables while omitting the other coefficients.[12]

[9] *ADH* also find that the absolute size of the import exposure coefficient rises in the IV specification.

[10] The Kleibergen-Paap statistic Kleibergen and Paap (2006,) is appropriate for use in the presence of non-i.i.d. errors, as opposed to the Cragg-Donald F statistic for the i.i.d. case.

[11] These critical values apply only to the i.i.d. case. Since there is no standard in how to test for weak instruments in the non-i.i.d. case, we follow Baum, Schaffer, and Stillman (2007) and use these critical values with some caution. Doing this appears to be more conservative than using the rule-of-thumb value of 10, suggested by Staiger and Stock (1997), which is only valid in the case of a single endogenous variable.

[12] The instruments are now constructed consistently from the import and export flows of the countries in the instrument group with Eastern Europe and, respectively, with China.

Table 2.2: Trade Exposure and Manufacturing Employment

	Dependent variable: 10-year change manufacturing employment / working age pop. in %-points					
	(1) OLS	(2) OLS	(3) OLS	(4) 2SLS	(5) 2SLS	(6) 2SLS
Δ import exposure	-0.047	-0.053	-0.068	-0.083	-0.154**	-0.177***
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)
Δ export exposure	0.352***	0.444***	0.418***	0.184	0.415**	0.387**
	(0.11)	(0.11)	(0.11)	(0.17)	(0.18)	(0.18)
% Manuf. of tradable goods	-0.079***	-0.091***	-0.087***	-0.054***	-0.078***	-0.073***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
% routine occupations		-0.073*	-0.072		-0.067*	-0.066
		(0.04)	(0.04)		(0.04)	(0.04)
% high skilled		-0.164***	-0.170***		-0.162***	-0.168***
		(0.04)	(0.04)		(0.04)	(0.04)
% foreigners		-0.061***	-0.059***		-0.060***	-0.059***
		(0.01)	(0.01)		(0.01)	(0.01)
% women		-0.038	-0.032		-0.031	-0.025
		(0.04)	(0.04)		(0.03)	(0.03)
Federal state dummies	Yes	Yes	-	Yes	Yes	-
Time dummy	Yes	Yes	-	Yes	Yes	-
State x time interactions	-	-	Yes	-	-	Yes
R-squared	0.338	0.477	0.496	0.192	0.365	0.264
First stage (KP)				20.232	18.294	17.203

Observations: 739. Standard errors clustered at spatial level (50 regions) in parentheses. All control variables are shares in total employment. % high skilled of labor force defined as the fraction of the workforce with a university degree. % routine occupations defined as basic activities according to Blossfeld (1987). Levels of significance: *** 1 %, ** 5 %, * 10 %.

Table (2.3) suggests that trade exposure with Eastern Europe had much stronger and more significant impacts on German manufacturing employment than trade exposure with China. For China, the coefficients are small and not (or only marginally) significant. For Eastern Europe, we find highly significant effects that are larger in absolute terms than the overall effects reported in Table (2.2). This suggests that the global effects of trade exposure with the East are actually driven by the import and export exposure with respect to Eastern Europe. A potential problem of this specification, however, is omitted variable bias since we consider trade exposure just with respect to one area while leaving out the (potentially relevant) exposure of the other area.

Net export exposure We tackle this issue in Table (2.4). Here we consider *net* export exposure of Germany with respect to China and Eastern Europe, which are now included in the same regression. For consistency, we instrument German net exposure with the net exports of the instrument countries vis-a-vis Eastern Europe and China, respectively. The message of Table (2.4) is consistent with our previous findings. The

Table 2.3: Trade exposure with Eastern Europe and China

	Dependent variable: 10-year change manufacturing employment / working age pop. in %-points					
	Eastern Europe trade			China trade		
Δ import exposure	-0.760*	-0.911**	-0.929**	-0.079	-0.121	-0.162*
	(0.44)	(0.40)	(0.37)	(0.07)	(0.07)	(0.08)
Δ export exposure	0.626*	0.905***	0.897***	-0.025	0.756	0.536
	(0.32)	(0.31)	(0.31)	(0.85)	(0.92)	(0.97)
Federal state dummies	Yes	Yes	-	Yes	Yes	-
Time dummy	Yes	Yes	-	Yes	Yes	-
State x time interactions	-	-	Yes	-	-	Yes
Further control variables	-	Yes	Yes	-	Yes	Yes
R-squared	0.155	0.287	0.166	0.157	0.376	0.261
First stage (KP)	12.697	12.482	13.227	11.983	10.528	10.268

Observations: 739. Standard errors clustered at spatial level (50 regions) in parentheses. IV estimates, including federal state and time interactions and all controls described in the benchmark specification. Levels of significance: *** 1 %, ** 5 %, * 10 %.

positive impact of export exposure seems to dominate the negative effect of import exposure on manufacturing employment in Germany. Furthermore, net export exposure only has a significant effect for Eastern Europe, but not for China, again suggesting that the impact of trade with the former area is economically more important for Germany.

Benchmarking the impact of trade on manufacturing employment What do these empirical results imply quantitatively? The results from Table (2.2) clearly suggest, that the rapid increase of trade integration with the East in the last 20 years had a positive overall effect and strengthened manufacturing employment in Germany. This can be seen from the higher estimated effect of exports relative to imports, and from the relatively stronger increase in export exposure relative to import penetration.

Our preferred estimates from column 6 of Table (2.2) imply that a 10-year change of € 1,000 per worker in import exposure reduces manufacturing employment relative to working age population by 0.177 percentage points in the aggregate, whereas export exposure increases this share by 0.387 percentage points. Taking into account that export exposure per worker increased by € 7,060 from 1988-2008 and import exposure by € 6,147, we can calculate that the new export opportunities increased normalized manufacturing employment by 2.73 percentage points. Import competition reduced it by "only" 1.09 percentage points, thus leading to a net increase in manufacturing employment in the economy as a result of the deeper trade integration.

To set these numbers into perspective, it is important to note that the manufacturing sector has been declining in Germany over the period 1988 to 2008 overall, representing a general trend of structural change away from manufacturing and towards modern

Table 2.4: Net Trade exposure and manufacturing employment

	Dependent variable: 10-year change manuf. emp. / working age pop. in %-points		
Δ net exposure to	0.671*	0.838**	0.825**
Eastern Europe trade	(0.40)	(0.38)	(0.38)
Δ net exposure to	-0.037	0.069	0.079
China trade	(0.14)	(0.13)	(0.13)
Federal state dummies	Yes	Yes	-
Time dummy	Yes	Yes	-
State x time interactions	-	-	Yes
Further control variables	-	Yes	Yes
R-squared	0.160	0.301	0.188
First stage (KP)	58.664	66.871	79.910

Observations: 739. Standard errors clustered at spatial level (50 regions) in parentheses. IV estimates, including federal state and time interactions and all controls described in the benchmark specification. Levels of significance: *** 1 %, ** 5 %, * 10 %.

service industries. Figure 2.2 shows that, in Western Germany, the share of manufacturing employment (measured in full-time equivalents) in the working age population dropped from 16 percent in 1988 to around 12 percent in 2008. This downward trend happened mostly in the first decade and then slowed down somewhat. Our estimates indicate that trade integration with Eastern Europe and China has slowed down this general trend, that is, it has retained manufacturing in the German economy in the past two decades. Below we conduct some additional quantitative explorations, where we benchmark the overall impact of trade on total employment in Germany (see Section 4.2.).

2.3.3 Robustness Checks

Identification

How robust are our results with respect to the definition of the "instrument group" of countries whose trade flows with China and Eastern Europe are used in the definition of (2.3) and (2.4)? To address this point, we re-estimate our baseline model with varying instruments (see Appendix-table 2.5).

In column 1, we first specify an over-identified model instead of the just identified IV model used as the benchmark. This approach exploits the detailed variation of the trade flows of the single instrument countries with China/Eastern Europe instead of adding up those trade flows. As can be seen, the results are similar as before, and the Hansen's J test which we can now perform further corroborates the validity of our instrument set.

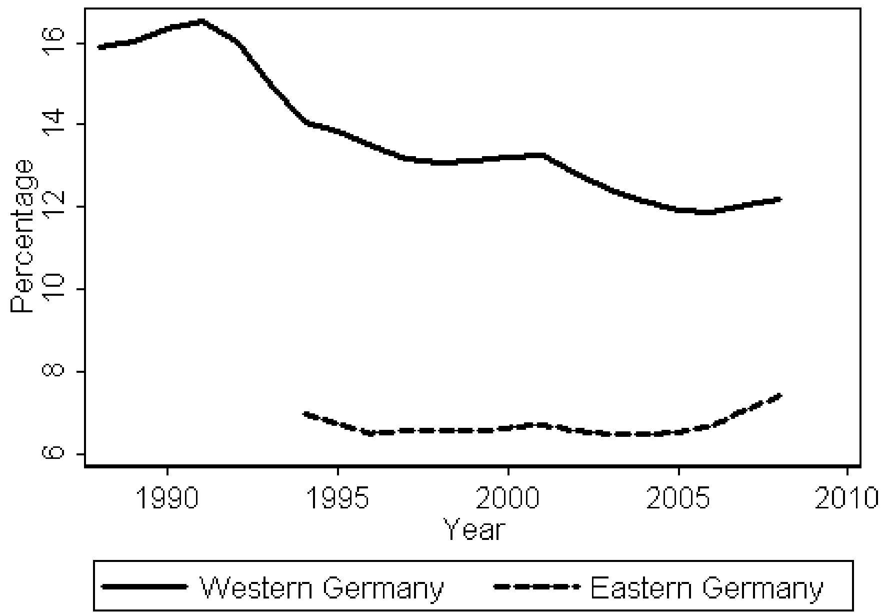


Figure 2.2: Percentage of manufacturing employees in working age population

In columns 2 to 4 we change the countries that are included in the instrument group. Recall that the validity of our identification approach hinges on the ability of the instrument to purge domestic shocks that simultaneously affect German regional employment and trade patterns. As explained above, we have therefore excluded direct neighbors of Germany as well as members of the European Monetary Union. There is still the concern that there might be an independent effect of the trade flows between China/Eastern Europe and those "instrument group" countries on German regions, which in turn would violate the exclusion restriction. This may be particularly relevant for the United Kingdom, which among the countries in the instrument group is the most important trading partner of Germany. We therefore drop the UK from the instrument group and re-estimate the (just identified) baseline specification. The results in column 2 show, however, that the estimation results are almost the same as in the baseline specification. In column 3 we add the USA to the instrument group, but again this hardly affects our estimation results. Finally, in column 4, we consider a placebo test by including only such countries in the instrument group, whose economic structures are totally dissimilar from Germany's, namely Cyprus, Iceland and the United Arab Emirates. As expected, the Kleibergen-Paap statistics indicate that these results are strongly biased due to weak instruments. Summing up, Table 2.5 suggests that our baseline specification indeed leads to a credible identification, as the adopted baseline instrument has both explanatory power in the first stage and does not violate the conditions for validity.

Another concern for identification is that the changes in manufacturing employment and trade exposure may be simultaneously driven by a common long-run trend. For example, employment in some manufacturing industries may have been on a secular decline even before the rise of the East kicked in, and the decreasing domestic production may then have been substituted by imports from the East. Similarly, industries may have boomed even before the mid-1980s, so that export exposure with the East was rather a symptom than a cause of domestic employment gains in manufacturing. The results in Appendix Table 2.6 suggest, however, that this is actually not the case. There we have considered a falsification test, where the change in manufacturing employment lagged by 10 years is regressed on the contemporaneous trade exposure with the East. The results show that lagged employment changes do not "predict" future regional trade exposure; in fact, coefficients are insignificant or even change sign. This robustness check thus further corroborates that our main results capture the *causal* effect of trade exposure on domestic manufacturing employment.

Particular industries

Next, we check the sensitivity of our results to the omission of specific industries. We re-estimate the baseline model and drop, in each specification, one industry from the data set which is among the top ten sectors when it comes to bilateral trade values in 2008 (Table 2.7 in the appendix). We find that leaving out the automobile industry or its most important suppliers (which constitute by far the most important export sector for the German economy) strongly decreases the coefficients for both import and export exposure. This highlights the importance of the car industry for both German manufacturing employment and trade. Omitting other industries, however, does not lead to a notable change in our estimated IV coefficients, compared to the baseline findings, although increasing standard errors sometimes render the estimated coefficients insignificant.

Regional classification

In the baseline specification, we have included all 413 (Eastern and Western) German regions in the analysis. Since we have data for Eastern Germany only after the German reunification, there are thus only 326 regions available in the first period. As a robustness check, we exclude all Eastern German regions also in the second period. The coefficients in Table 2.8 (columns 1–3) in the appendix are similar as in our baseline estimation, so that all conclusions are qualitatively unchanged.

Finally, we investigate the robustness of our results with respect to the regional level of analysis. As an alternative to the 413 administrative NUTS-3 regions, we consider 50 aggregate labor market regions (Kropp and Schwengler (2011)), which are comparable constructs to the US commuting zones used by *ADH*. The resulting coefficients in columns 4–6 of Table 2.8 are also similar to our baseline specification, though standard errors are larger. We thus prefer to stick to the more detailed regional level that offers more heterogeneity.

2.4 Other Regional Labor Market Outcomes

In this section we consider the impact of the rise of the East on other labor market outcomes across German regions.

2.4.1 Population Shifts

The first important question is whether trade exposure induces population shifts across regions. In fact, if labor were perfectly mobile across space, workers should respond instantaneously to trade shocks by relocating between regions. The differential response of employment across local labor markets would then be less informative about the effects of trade liberalization, while the impacts would become visible in regional migration patterns or adjustments of local population sizes. In their analysis on the impact of Chinese import exposure, *ADH* emphasize that there seems to be a sluggish adjustment of population across local labor markets in the US. That is, labor markets seem to have adjusted mainly at the employment margin while there have been little population shifts in response to the (potential) Chinese import competition. In this subsection we analyze if a similar pattern emerges in the German case. Moreover, recall that the main outcome variable in the analysis above is the share of regional manufacturing employment in the total working age populations. To disentangle the impact of trade exposure on this outcome variable, it is therefore important to study the effects on regional population shifts.

The estimation results are reported in Table 2.5, column 1.[13] As can be seen, the impact of overall export exposure on the 10-year change in (log) regional working age populations is statistically not different from zero. That is, regions with industrial structures more strongly exposed to the new export opportunities in the East did not experience significant inward migrations, or other forms of population gains. For import exposure, we find a slightly negative impact on regional population sizes. This impact is weak at best, however, and significant only at the 10% level.

These findings, in combination with our baseline results from above, thus suggest that the adjustment in the German labor markets occurred mainly at the employment margin, that is, via the creation or destruction of manufacturing jobs, while there have

[13] All specifications in Table 2.5 are analogous to the baseline IV regression from column 6 of Table 2.2. For brevity we again focus on the main variables and omit the coefficients for the other controls.

Table 2.5: Other labor market indicators

	Dependent variables: 10-year change			
	log working age population	Total emp.	Unemployment / working age pop. in %-points	Non-manuf. emp.
Δ import exposure	-0.242*	-0.333**	0.005	-0.156
	(0.14)	(0.11)	(0.02)	(0.11)
Δ export exposure	0.244	0.663**	-0.097**	0.276
	(0.19)	(0.27)	(0.04)	(0.19)
R-square	0.151	0.103	0.070	0.177

Observations: 739. Standard errors clustered at spatial level (50 regions) in parentheses. IV estimates, including federal state and time interactions and all controls described in the benchmark specification. Levels of significance: *** 1 %, ** 5 %, * 10 %.

been little or no induced population shifts. This interpretation is also consistent with the results reported in Appendix Table 2.9, where we re-estimate the baseline specification from above using the log change in *absolute* regional manufacturing employment, not divided by regional working age population, as the outcome variable. We obtain coefficients that are qualitatively in line with our main results.

Our finding that trade exposure has mainly affected employment rather than population sizes in German regions is in line with *ADH*'s results for the US case, which is plausible since it has been frequently argued in the literature (Molloy, Smith, and Wozniak (2011)) that regional labor mobility is even lower in Germany than in the US .

2.4.2 Total Regional Employment and Unemployment

Columns 2 and 3 of Table 2.5 show that higher export exposure raises *total* regional employment, again measured relative to working age population, and lowers regional unemployment. However, non-manufacturing employment is not significantly positively affected by export exposure as is shown in column 4. That is, the rise of the East seems to benefit regions with export oriented industrial structures mainly through additional manufacturing jobs, which in turn raises the overall regional employment rate and reduces unemployment. Local "spillovers" of export exposure to the non-manufacturing sector may exist, for example, through a higher demand for services from the expanding manufacturing sector, since the impact of export exposure on non-manufacturing employment is estimated to be positive. Yet, standard errors are fairly large so that evidence does not generally support the hypothesis that export opportunities in the East also generate jobs beyond the tradable goods sector. We return to this issue in the next subsection, where we further disentangle employment reactions in different non-manufacturing industries that are not directly affected by the new market opportunities in the East, but that may be indirectly affected.

Turning to the impact of import exposure, we obtain results that largely mirror these effects. Regions with industrial structures more strongly exposed to import competition saw a stronger decline not only in manufacturing employment, but also in the total employment rate. Non-manufacturing employment also seems to be negatively affected, but the respective coefficient is again not significant. In short, import penetration from the East has caused job losses, clearly so in the manufacturing sector and possibly beyond. However, one dimension along which the results for import and export exposure seem to differ, is that a higher import exposure apparently does not increase regional unemployment. The estimated coefficient is positive, but it is fairly small and statistically insignificant. There are two possible explanations for this finding. First, in Germany, there are numerous active labor market policies that target workers who have been laid off (or face a substantial risk thereof). These programs may cushion possible adverse import shocks, as workers prone to becoming unemployed are either retained in their original job via measures such as *Kurzarbeit* where they reduce hours but remain with their original firm, or they may be transferred into an active labor market measure fairly quickly, in which case they are not counted as unemployed. Second, recall that we have found at least a small impact of import exposure on population shifts (see column 1 of Table 2.5), which suggests that at least some workers respond to local import shocks with migration to other regions with more favorable industrial structures.

Benchmarking the impact of trade Summing up, trade exposure seems to have broad employment effects on the affected regions such that export oriented regions experienced a net gain from the rise of the East, while import competing regions faced comprehensive job losses. Multiplying the coefficients from column 2 of Table 2.5 with the average observed increase in trade exposure per worker, we can calculate that export exposure increased total employment over working age population in the average region by 4.68 percentage points, while import exposure lowered it by 2.05 points. This suggests that there is a sizeable positive net impact of the rise of the East on total employment in Germany, somewhere in the ballpark of 1 million additional jobs that were created between 1988 and 2008 as a result of trade.

However, as we argued above, we employ our IV strategy to recover the causal effect of export and import exposure across local labor markets. Still, the exposure variables as constructed in (2.3) and (2.4) may contain German supply and demand shocks in addition to the exogenous component, namely the rise of the East in the world economy. Our back-of-the-envelope calculations are, hence, likely to overstate the effect of trade integration on normalized employment changes. To address this, we follow *ADH* and employ a simple decomposition exercise. The idea is to isolate the share of the exposure variables (2.3) and (2.4), which is driven by the exogenous forces of increased trade exposure.[14] This gives a more conservative estimate of the impact of exports on employment over working age population of 2.34 percentage points. Analogously, this procedure yields an estimate of -1.38 percentage points for the impact of imports. These estimates together imply a gain of 492,455 full-time equivalent jobs in the period 1988-2008 that would not exist without the rise of the East. Notice that this is a *net* gain of jobs over a twenty-year period, brought about by an aggregate increase in the employment rate in the German economy.

2.4.3 Disentangling the Impact of Non-Manufacturing Industries

In the last step of the aggregate analysis, we investigate in greater detail the impact of trade exposure on employment in non-manufacturing industries. Recall that we have not found statistically significant effects when lumping all non-manufacturing activities together (see column 4 of Table 2.5). However, those coefficients for the overall effects may mask more specific impacts of trade on particular industries within that category. In Table 2.10 in the Appendix we distinguish four different non-manufacturing sectors (construction, personal services, business services and the public sector) and re-estimate our baseline specification for each of those industry groups.

As can be seen, there are virtually no effects on local employment in construction or personal services, neither with respect to import nor with respect to export exposure. However, we do find sizable and statistically highly significant employment effects in business service industries that go into the same direction as the employment effects in the manufacturing sector. More specifically, a region strongly exposed to exports to the East not only experienced job gains within the tradable goods sector (manufacturing) but also in local business services. The reason can be a localized cross-industry demand spillover: As manufacturing industries expand in Eastern markets, they not only build up domestic employment in the own industry but also require further intermediate inputs such as business services. The induced demand generates jobs in those service industries, and this effect seems to be locally tied to the rise of the downstream manufacturing sector. Analogously, regions with higher import exposure experienced stronger job losses not only in the manufacturing sector that is directly affected by the displacement from Eastern import penetration, but also suffered from complimentary job losses in business services. For personal services and construction, we do not find evidence for such spillovers of trade on employment, at least these spillover effects do not appear to be localized in the German case.

[14]The decomposition relies on the relationship between the IV and the OLS estimators. See *ADH* for details. Performing the exercise separately for exports and imports, we estimate that the fraction in the export exposure variables that is explained by exogenous forces to be 0.499 and 0.675 for imports.

As for the impact of trade exposure on local public sector employment, we find that it is also virtually nil. On the one hand, demand spillovers from manufacturing to the public sector are very unlikely to play a role, which is consistent with our empirical findings. Yet, the government may try to compensate job losses in private industry by expanding public employment particularly in adversely affected locations (Faggio and Overman (2012)), such as locations that face stiff import penetration. However, for the case of Germany we do not find evidence for such an effect of trade on public sector jobs.

2.5 Worker Level Evidence

The analysis so far has focussed on the impact of trade exposure on regional labor market aggregates. In this final section, we extend our analysis along the lines of Autor, Dorn, Hanson, and Song (2012) to the individual level, using detailed micro data on employment histories of German manufacturing workers.

From the perspective of a single worker, trade liberalization may increase the risk of displacement, if the own job is subject to high (potential) import competition. An extensive literature (Topel (1990), von Wachter and Bender (2006), Sullivan and von Wachter (2009)) documents that if displaced workers have to find new jobs and acquire human capital specific to their new employers, this in turn can lead to adverse effects on employment biographies in terms of reduced employment and earnings spells. On the other hand, export opportunities can have a countervailing stabilizing effect on individual employment relationships. Workers who are involved in the production of goods that are increasingly in demand from abroad, might face a lower probability of job termination. Holding everything else constant, they may even be able to accumulate firm- and industry-specific human capital and raise their long-term labor market prospects.

2.5.1 Data and Variables

We use the Sample of Integrated Labour Market Biographies (SIAB, Dorner, Heining, Jacobebbinghaus, and Seth (2010)). This data stems from all German social security notifications in the years 1975 to 2008. A two percent random sample has been drawn from all persons who have either been employed or officially registered as job-seekers resulting in an individual-level spell data set with information on age, sex, nationality, qualification, occupation, spell durations, etc. This data is highly accurate even on a daily base due to its original purpose of calculating retirement pensions. Since the notifications of employees are passed by their employers, establishment level data from the Establishment History Panel (BHP) can be merged to this data set. To be consistent with the periods considered at the regional level, we analyze individuals who have been employed in the manufacturing sector either in 1988 or 1998 and construct our dependent variable as cumulative days in employment over the following ten years. We only consider working age persons (22 – 64 years) in the respective period.

The trade exposure indices are constructed similarly as before. Yet, we now construct them at the *industry* level, in order to measure trade exposure at the level of an individual worker. The intuition is that manufacturing workers often have acquired sector- and occupation-specific human capital, so that they cannot switch instantaneously between occupations and industries. The change in import penetration per worker from both China and Eastern Europe (indexed by k) over the period $t = \{1988 - 1998, 1998 - 2008\}$ in a German industry j is defined as

$$\Delta IP_{jt} = \frac{\Delta Im_{jt}^{EAST}}{E_{jt}}, \tag{2.6}$$

where ΔIm_{jt}^{EAST} is the change in imports from China and Eastern Europe to Germany over period t, and E_{jt} is total employment in industry j at the beginning of the period. Analogously, the change in export opportunities per worker in industry j is

$$\Delta EP_{jt} = \frac{\Delta Ex_{jt}^{EAST}}{E_{jt}}, \tag{2.7}$$

where ΔEx_{jt}^{EAST} is the respective change in exports of industry j from Germany to China and Eastern Europe. See Table 2.6 for an overview of the data.

Our focus is the identification of the causal effect of the rise of the East on individual worker biographies in German manufacturing. Hence, we again rely on a instrumental variable approach for identification. We construct the following instruments:

$$\Delta IP_{ijt} = \frac{\Delta Im_{j-3t}^{EAST-other}}{E_{j-3t-3}} \qquad \text{and} \qquad \Delta EP_{ijt} = \frac{\Delta Ex_{j-3t}^{EAST-other}}{E_{j-3t-3}} \tag{2.8}$$

where we use the trade flows of the same set of countries as in the previous section. We use lagged employment shares of the sectors where workers were employed three years prior to the start of the period to avoid a possible influence of sorting of workers due to anticipation of future trade exposure.

In the regression, we again control for the regional shares of tradeable goods industries and interaction terms for federal states and time periods. Additionally, we use standard Mincerian individual-level variables in the list of controls, as well as dummies to control for year of birth. Since import and export exposure only vary across industries, one might worry that they capture industry-level effects that correlate with the change in trade exposure. To mitigate this multi-level problem, we also include further industry-level control variables in the regression, more specifically the Herfindahl-Index of establishment sizes, the Ellison and Glaeser (1997) agglomeration-index, the share of plants younger than two years, the average establishment size, the share of highly qualified employees, and the share of employees older than 50. Throughout, we allow our standard errors to be correlated between workers within the same industry and federal state.

Table 2.6: Means and standard deviations of main variables for manufacturing workers

	1988-1998		1998-2008	
	Outcome variables			
Cumulative years of employment	7.50	(3.03)	7.85	(2.96)
Cumulative years of employment in original establishment	5.68	(3.72)	5.58	(3.90)
Cumulative years of employment in original 3-digit industry	6.10	(3.67)	6.21	(3.82)
Cumulative years of employment in original labor market region	7.04	(3.28)	7.17	(3.39)
	Trade exposure			
Δ imports per worker$_{t=0}$				
Eastern Europe	4.74	(4.92)	6.61	(9.42)
China	1.55	(3.85)	6.60	(20.26)
Both	6.32	(7.25)	13.24	(22.80)
Δ exports per worker$_{t=0}$				
Eastern Europe	5.92	(5.54)	13.16	(10.81)
China	0.39	(0.96)	3.86	(4.40)
Both	6.29	(5.93)	17.33	(13.44)

Trade exposure measured in € 1,000 per worker

2.5.2 Results

The first two columns in Table 2.7 display the effects of an increase in Eastern trade exposure on the total number of days in employment over a 10 year period. While column (1) refers to the OLS estimation, we implement our IV strategy in column 2. The interpretation of the export exposure coefficient in column 2 is that a € 1,000 increase in industry exports per worker increases the expected time of employment over 10 years by 3.32 days ($= 0.91 \cdot \frac{365}{100}$), ceteris paribus. Given that the average worker in manufacturing has faced an increase of export exposure by more than € 17,000 over a ten year period, this implies that expected employment at the worker level has increased by about 56 days due to increasing export exposure. At the same time, an increase in import exposure has an opposing negative effect on job stability. For a worker who faces the average increase in imports by €6,290 in the second period, we estimate that time of employment over 10 years is reduced by 8.3 days. These results imply that the rise of the East overall has stabilized employment relationships and reduced the individual risk of job termination. This confirms our previous findings at the regional level, namely that exports opportunities on average more than offset the negative effects of rising import competition from the East.

Table 2.7: Eastern trade exposure and individual employment

	Dependent variable: 100 x cumulative years of employment over 10 year period				
	OLS	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)
	total	total	plant	3-digit ind.	region
Δ Imports	-0.17**	-0.36***	-1.04***	-0.92***	-0.61***
per worker$_{t=0}$	(0.08)	(0.12)	(0.22)	(0.21)	(0.15)
Δ Exports	0.85***	0.91***	1.36*	1.11*	1.46***
per worker$_{t=0}$	(0.14)	(0.27)	(0.78)	(0.63)	(0.32)
Employment in a tradable	5.23	5.33	14.93*	-18.43**	9.11*
goods industry in $t = 0$	(4.06)	(4.17)	(8.32)	(7.48)	(4.86)
Female	-182.27***	-181.99***	-127.36***	-146.56***	-160.01***
	(2.89)	(2.84)	(3.64)	(3.50)	(2.93)
Foreign citizen	-52.78***	-52.70***	-27.94***	-36.41***	-39.69***
	(2.68)	(2.69)	(3.35)	(3.45)	(2.98)
Low skilled	-29.25***	-29.02***	-16.19***	-21.98***	-17.86***
	(1.88)	(1.87)	(2.90)	(2.66)	(2.19)
High skilled	32.90***	32.99***	-43.89***	-23.64***	-32.74***
	(3.27)	(3.25)	(5.78)	(5.87)	(5.40)
Industry level controls	Yes	Yes	Yes	Yes	Yes
R-square	0.197	0.112	0.085	0.087	0.086
First Stage (KP)		23.155	23.155	23.155	23.155

Observations: 185,335. Standard errors clustered by 1,279 industry × federal state cells in parentheses. Control variables include dummy variables for start of period tenure, plant size, year of birth and federal state × period fixed effects. Models (3) – (5) consider cumulative employment only within the original establishment, 3-digit industry, and region, respectively. * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$

Our data permits us to further disaggregate the effect, and to investigate how trade exposure affects job stability for individual workers at the plant-, industry-, or region-level. Such effects might not be visible when looking only at total employment, since individuals might have changed jobs across plants, industries, or regions without a notable unemployment spell. The results reported in columns 3–5 indeed show that trade exposure with the East has caused significant job turnover that is not observable at the aggregate level. Increased exposure to import competition by €1,000 reduces the expected time spent with the original employer by 3.8 days and, respectively, the original 3-digit industry by 3.4 days. That is, import exposure has causally increased job churning both within and across industries. On the other hand, rising export exposure has a converse but less precisely estimated effect on those job stability indicators. Furthermore, we find that employees in industries with high export exposure are significantly less likely to relocate to other regions. These findings are in line with and complementary to the aggregate results discussed before.

2.6 Discussion and Conclusion

The past decades have seen a strong increase in the volume of international trade. Deregulation and the abolishment of trade barriers as well as drastic reductions in transport costs have led to a steadily increasing integration of national economies. In this paper, we focus on two major facets of globalization: China's explosive ascent and the rise of Eastern Europe after the fall of the Iron Curtain. Understanding the consequences of those developments for the labor markets in the traditional Western market economies is crucial, both from an economic and a political point of view.

We analyze the *causal* impact of the rise of China and Eastern Europe on the performance of local labor markets in Germany during the period 1988 to 2008, using an instrumental variable approach pioneered by Autor, Dorn, and Hanson (2012). At the regional level, Germany is characterized by a substantial variation in local industrial structures. These initial structures determine how the regions were affected by the rising trade exposure that kicked in since the mid 1990s.

Two main messages can be derived from our analysis: First, the rise of Eastern Europe had much more immediate consequences for the German economy than the rise of China. Second, overall, the rise in trade exposure has led to substantial employment gains in the German economy, but these gains are highly unevenly distributed across space. In fact, some regions have lost jobs as a result of the deeper trade integration, both in the manufacturing sector and beyond. But those losses were, in the aggregate, more than offset by additional jobs created in regions with industrial structures that allowed them to take advantage of the new export opportunities in the East. In our analysis at the individual level we complement this aggregate picture and show that trade exposure has, overall, led to more stable employment relationships by reducing the risk of job termination. However, trade again produces winners and losers, since workers in import competing industries indeed faced an increased risk of job churning and lower overall employment spells.

Our results for the German economy differ quite substantially from the findings of Autor, Dorn, and Hanson (2012) for the United States. Trade liberalization with China is likely to bring about welfare gains also for the US case, for example through gains in productivity or consumption diversity. Yet, these authors stress that in the short-to-medium run, the US economy has to face severe adverse effects on local labor markets, even when taking into account that the rise of China not only creates import penetration but also new export opportunities. The situation of Germany seems to be quite different, at least on average, as the overall labor market consequences are largely positive even in the medium run. This finding may be explained by the fact that overall trade with China is much more balanced in the German than in the US case. Furthermore, our analysis suggests that focusing only on China provides an incomplete picture. The rise of Eastern Europe had a much stronger impact on German local labor markets than the rise of China, possibly reflecting the fact that the Eastern European markets are located (much) closer by.

Germany might be a special case due to its large trade surplus. Yet, America's extreme trade deficit is also very unique and only sustainable because of its status as the World's largest economy. Smaller industrialized countries, such as France, Italy, South Korea, or Japan, that are not able to sustain a trade deficit of this magnitude might be more comparable to Germany than to the US in the long run. Our results suggest that the experience for these countries may be closer to that of Germany, as long as their economies similarly retain a focus on modern, export oriented manufacturing.

In our main analysis, we assign sector level trade data to German regions according to their initial industrial structures. This approach has the caveat that we can only observe the *potential* trade exposure with the East. It is not possible to directly relate trade flows to specific firms or local industries. Hence, we have to assume that all German firms in a given sector are affected more or less uniformly by the rise of the East. An advantage of our approach is that it allows to analyze the local adjustments to trade exposure along many different margins. Our main focus on manufacturing employment is interesting, because in most industrialized countries there has been a long-run trend of structural change where employment secularly shifted away from the manufacturing sector and towards modern service industries. Our results suggest that trade with the East has per se decelerated this declining trend, and contributed to retaining the manufacturing sector in the German economy.

2.1 Appendix – Chapter Two

2.1.1 Tables and Graphs

Table 2.1: Means and standard deviations of main variables

	1988-1998		1998-2008	
	Outcome variables			
10-year change manuf. employment / working age pop. in %-points	-2.51	(2.71)	-0.15	(2.21)
	Trade exposure			
Change in import exposure				
Eastern Europe	1.80	(1.00)	1.85	(1.30)
China	0.59	(0.52)	1.90	(1.88)
Both	2.40	(1.32)	3.75	(2.65)
Change in export exposure				
Eastern Europe	2.17	(1.01)	3.71	(2.27)
China	0.13	(0.11)	1.04	(0.82)
Both	2.31	(1.05)	4.75	(3.00)
	Control variables			
Initial shares in total labor force				
Manuf. of tradable goods	35.52	(12.81)	27.42	(12.69)
Routine occupations	41.34	(4.46)	36.42	(4.41)
High skilled	4.30	(2.43)	7.09	(3.76)
Foreigners	6.46	(3.71)	5.86	(4.26)
Women	38.50	(13.98)	40.41	(13.35)

Trade exposure in € 1,000 per worker. Control variables in percent.

The sectoral composition of German trade

Table 2.2: Trade volumes of the top ten sectors in trade with Eastern Europe

	Industry	2008	1998	1988
Imports from Eastern Europe				
111	Extraction of crude petroleum and natural gas*	20700	2340	1460
341	Manuf. of motor vehicles	7100	4440	76
343	Manuf. of parts and accessories for motor vehicles and their engines	6830	1610	11
274	Manuf. of basic precious and non-ferrous metals	4280	1940	992
271	Manuf. of basic iron and steel and of ferro-alloys (ECSC1)	3510	949	402
316	Manuf. of electrical equipment n.e.c.	3350	1260	26
361	Manuf. of furniture	3260	2260	449
291	Manuf. of machinery for the production and use of mechanical power, except aircraft, vehicle and cycle engines	3080	727	85
241	Manuf. of basic chemicals	3010	1300	442
287	Manuf. of other fabricated metal products	2500	1190	75
Exports to Eastern Europe				
341	Manuf. of motor vehicles	13300	3970	248
343	Manuf. of parts and accessories for motor vehicles and their engines	9180	2610	92
295	Manuf. of other special purpose machinery	7830	3400	1250
291	Manuf. of machinery for the production and use of mechanical power, except aircraft, vehicle and cycle engines	5390	1500	413
252	Manuf. of plastic products	5280	2090	577
241	Manuf. of basic chemicals	4990	1540	989
292	Manuf. of other general purpose machinery	4500	1710	447
287	Manuf. of other fabricated metal products	4030	1360	128
244	Manuf. of pharmaceuticals, medicinal chemicals and botanical products	3950	1000	245
312	Manuf. of electricity distribution and control apparatus	3900	1440	155

Trade volumes measured in Million Euros of 2005. *: This industry and all other industries related to agriculture, mining and fuel products are omitted in the empirical analysis.

Table 2.3: Trade volumes of the top ten sectors in trade with China

	Industry	2008	1998	1988
Imports from China				
300	Manuf. of office machinery and computers	8630	1160	12
182	Manuf. of other wearing apparel and accessories	4950	1900	704
365	Manuf. of games and toys	3280	658	46
323	Manuf. of television and radio receivers, sound or video recording or reproducing apparatus and associated goods	2930	700	171
321	Manuf. of electronic valves and tubes and other electronic components	2920	123	2
322	Manuf. of television and radio transmitters and apparatus for line telephony and line telegraphy	1740	172	8
287	Manuf. of other fabricated metal products	1510	390	40
177	Manuf. of knitted and crocheted articles	1360	199	24
241	Manuf. of basic chemicals	1200	335	115
297	Manuf. of domestic appliances n.e.c.	1190	392	10
Exports to China				
341	Manuf. of motor vehicles	3530	238	209
295	Manuf. of other special purpose machinery	3220	1050	590
291	Manuf. of machinery for the production and use of mechanical power, except aircraft, vehicle and cycle engines	2740	248	108
294	Manuf. of machine-tools	1900	376	306
312	Manuf. of electricity distribution and control apparatus	1650	277	54
343	Manuf. of parts and accessories for motor vehicles and their engines	1640	114	31
292	Manuf. of other general purpose machinery	1570	388	112
353	Manuf. of aircraft and spacecraft	1310	182	11
332	Manuf. of instruments and appliances for measuring, checking, testing, nav. and other purposes, except industrial process control equipment	1220	168	84
311	Manuf. of electric motors, generators and transformers	1200	83	26

Trade volumes measured in Million Euros of 2005.

Further results

Table 2.4: First stage regressions

	Dependent variables:					
	Δ import exposure			Δ export exposure		
Δ import exposure (inst)	0.247***	0.247***	0.248***	-0.006	-0.006	-0.009
	(4.28)	(4.31)	(4.18)	(-0.41)	(-0.41)	(-0.61)
Δ export exposure	0.124***	0.119***	0.116***	0.422***	0.423***	0.415***
	(4.11)	(3.68)	(3.71)	(7.12)	(6.84)	(6.54)
Further control variables	No	Yes	Yes	No	Yes	Yes
Federal state dummies	Yes	Yes	No	Yes	Yes	No
Time dummy	Yes	Yes	No	Yes	Yes	No
State x time interactions	No	No	Yes	No	No	Yes
R-squared	0.764	0.766	0.771	0.822	0.822	0.827
F-test of excluded inst.	19.39***	17.33***	15.18***	27.42***	25.54***	22.71***

Observations: 739. OLS estimates, t-values in parentheses, based on standard errors clustered at spatial level (50 regions). Levels of significance: *** 1 %, ** 5 %, * 10 %.

Table 2.5: Robustness checks: Variations in instrumental variables

	Dependent variable: 10-year change manufacturing employment / working age pop. in %-points			
	Over-identified	Leave out UK	Add USA	Placebo CY, IS, UAE
Δ import exposure	-0.116**	-0.175***	-0.188**	-0.124
	(0.05)	(0.07)	(0.08)	(0.13)
Δ export exposure	0.377**	0.385**	0.362*	0.282
	(0.15)	(0.19)	(0.19)	(0.25)
R-squared	0.269	0.264	0.261	0.260
First stage (KP)	58.993	12.607	18.623	3.659
p Hansen	0.113			

Observations: 739. Standard errors clustered at spatial level (50 regions) in parentheses. IV estimates, including federal state and time interactions and all controls described in the benchmark specification. Levels of significance: *** 1 %, ** 5 %, * 10 %.

Table 2.6: Falsification: Lagged change in manuf. employment and future trade exposure

	Dependent variable: Lagged 10-year change manufacturing employment / working age pop. in %-points					
	Joint		Eastern Europe trade		China trade	
Δ import exposure	0.080	0.105*	0.542**	0.482*	-0.131	-0.053
	(0.06)	(0.05)	(0.24)	(0.26)	(0.08)	(0.06)
Δ export exposure	-0.064	0.005	-0.152	-0.058	-0.881*	-0.512
	(0.12)	(0.11)	(0.16)	(0.16)	(0.45)	(0.49)
Lagged control vars.	-	Yes	-	Yes	-	Yes
R-squared	0.196	0.355	0.215	0.370	0.223	0.359

Observations: 652. Standard errors clustered at spatial level (50 regions) in parentheses. OLS estimates, including federal state and time interactions and all controls described in the benchmark specification. Levels of significance: *** 1 %, ** 5 %, * 10 %.

Table 2.7: Robustness checks: Drop most important industries for trade with Eastern Europe and China

	Dependent variable: 10-year change manufacturing employment / working age pop. in %-points				
Omitted industry	Motor vehicles	Spec. purp. machinery	Parts for motor vehicles	Mach. for prod. of mech. power	Basic chemicals
Δ import exposure	-0.110	-0.171***	-0.112	-0.149**	-0.166***
	(0.08)	(0.07)	(0.07)	(0.07)	(0.06)
Δ export exposure	0.141	0.354	0.220	0.180	0.616***
	(0.29)	(0.23)	(0.23)	(0.22)	(0.20)
R-square	0.273	0.255	0.275	0.240	0.281
First Stage (KP)	8.288	14.894	15.936	14.294	11.328
Omitted industry	Office machines	Wearing apparel	Communication devices	Electrical equipment	Furniture
Δ import exposure	-0.114	-0.136**	-0.121*	-0.120*	-0.082
	(0.09)	(0.06)	(0.07)	(0.06)	(0.07)
Δ export exposure	0.353	0.305	0.281	0.338	0.253
	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)
R-square	0.275	0.267	0.261	0.283	0.279
First Stage (KP)	13.144	15.380	15.077	15.715	16.280

Observations: 739. Standard errors clustered by administrative districts and years in parentheses. IV estimates, including federal state and time interactions and all control variables. * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$

Table 2.8: Robustness checks: Regional Classification

	Dependent variable: 10-year change manufacturing employment / working age pop. in %-points					
	(1)	(2)	(3)	(4)	(5)	(6)
	Western Germany only			50 labor market regions		
Δ import exposure	-0.124*	-0.204***	-0.229***	-0.170	-0.004	-0.092
	(0.06)	(0.07)	(0.07)	(0.15)	(0.19)	(0.18)
Δ export exposure	0.217	0.446**	0.418**	0.367***	0.296*	0.321**
	(0.17)	(0.18)	(0.18)	(0.13)	(0.17)	(0.15)
% Manuf. of tradable goods	-0.059***	-0.081***	-0.075***	-0.065**	-0.009	-0.008
	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)
Further controls	-	Yes	Yes	-	Yes	Yes
Federal state dummies	Yes	Yes	-	Yes	Yes	-
Time dummy	Yes	Yes	-	Yes	Yes	-
State and time interaction	-	-	Yes	-	-	Yes
R-squared	0.212	0.389	0.287	0.530	0.692	0.457
First stage (KP)	20.523	18.347	16.886	54.511	26.832	20.803

Observations: 652/100. Standard errors clustered by labor market regions in parentheses. * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$

Table 2.9: Alternative definition of dependent variable

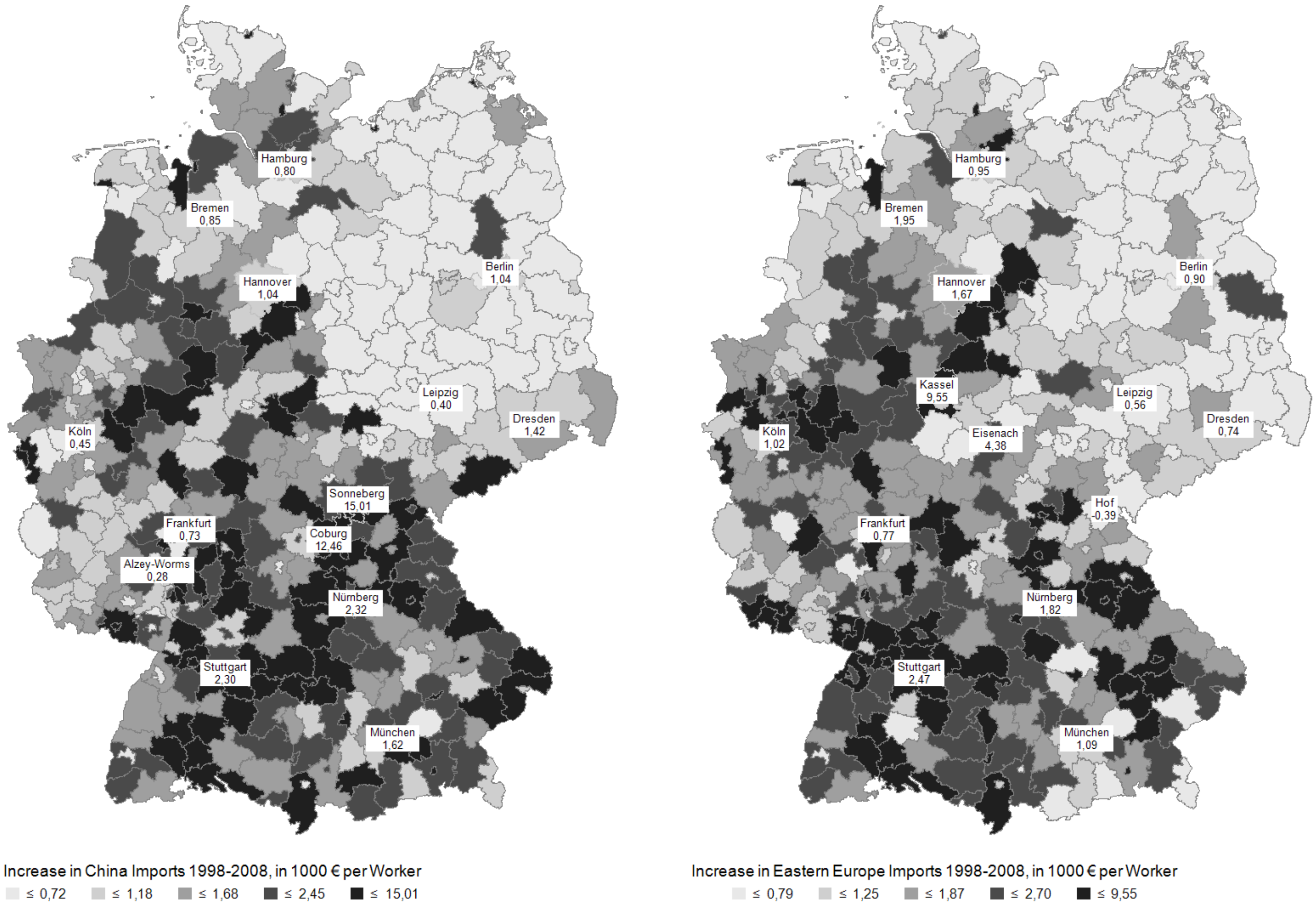
	Dependent variable: 10-year change ln employment in manufacturing	ln employment in non-manufacturing	ln employment in total	ln un-employment
Δ import exposure	-1.480***	-0.761*	-0.945**	-0.196
	(0.54)	(0.45)	(0.43)	(0.44)
Δ export exposure	1.638**	1.287**	1.175***	-1.237*
	(0.66)	(0.61)	(0.39)	(0.66)
R-squared	0.164	0.148	0.155	0.045
First stage (KP)	17.203	17.203	17.203	17.203

Observations: 739. Standard errors clustered by 50 labor market regions in parentheses. Coefficients and standard errors multiplied times 100. IV estimates, including federal state and time interactions and all controls described in the benchmark specification. Levels of significance: *** 1 %, ** 5 %, * 10 %.

Table 2.10: Impact on non-manufacturing industries

	Dependent variables: 10-year change in employment / working age pop. in %-points			
	cons-truction	personal services	business services	public sector
Δ import exposure	0.011	-0.056	-0.101*	-0.014
	(0.01)	(0.04)	(0.06)	(0.02)
Δ export exposure	0.021	-0.000	0.260***	-0.007
	(0.02)	(0.06)	(0.10)	(0.02)
R-squared	0.159	0.113	0.396	0.095
First stage (KP)	17.203	17.203	17.203	17.203

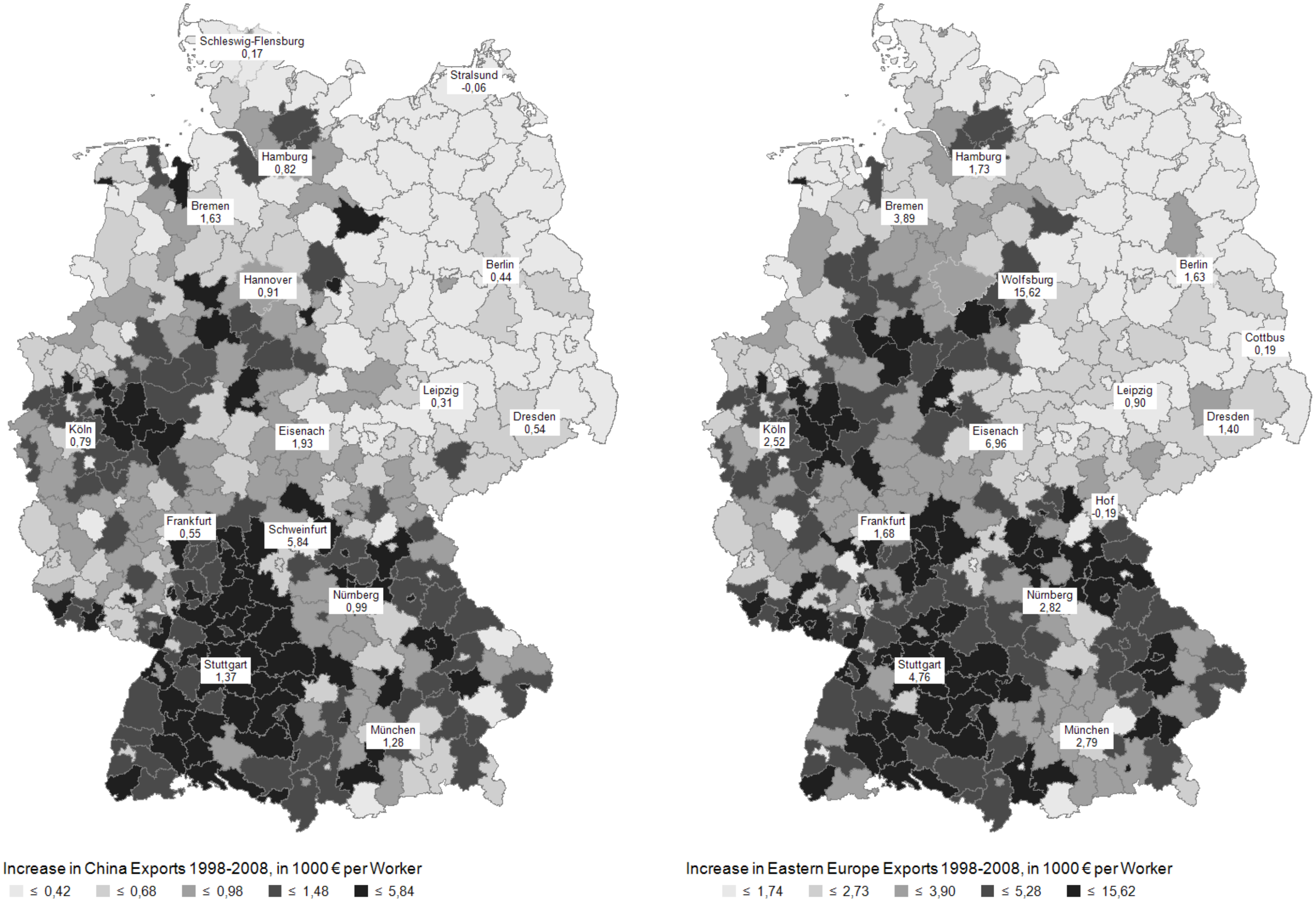
Observations: 739. Standard errors clustered by 50 labor market regions in parentheses. IV estimates, including federal state and time interactions and all controls described in the benchmark specification. Levels of significance: *** 1 %, ** 5 %, * 10 %.



(a) China

(b) Eastern Europe

Figure 2.1: Change in import exposure 1998-2008



(a) China

(b) Eastern Europe

Figure 2.2: Change in export exposure 1998-2008

Chapter 3

Efficient Labor and Capital Income Taxation over the Life Cycle[1]

This chapter is joint work with Dominik Sachs.

3.1 Introduction

How can a government efficiently redistribute income among individuals? Should the government solely rely on the taxation of labor income or should capital be taxed as well? What role does the evolution of inequality over the life cycle play for the design of optimal policies?

In this paper we address these questions in a life cycle framework. Consistent with a large empirical literature, inequality changes over the life cycle, both because of forecastable heterogeneity across individuals and because of idiosyncratic risk. Consistent with current tax practices, taxes condition on current (annual) earnings. The labor income tax is allowed to be fully non-linear in the tradition of the seminal approach to optimal taxation pioneered by Mirrlees (1971). The tax on wealth (or equivalently capital income) is linear.

[1]Contact: sebastian.findeisen@uzh.ch, dominik.sachs@uni-konstanz.de. We are grateful to Emmanuel Saez for valuable comments. We thank Alan Auerbach, Leo Kaas and seminar participants in Berkeley and Konstanz for discussions. Sebastian Findeisen gratefully acknowledges the hospitality of Berkeley. We are thankful to Fatih Karahan and Serdar Ozkan who kindly made their estimates available to us.

We show that a novel and simple formula for the optimal taxation of capital arises in our simple life cycle model. It follows a standard public finance trade-off: the tax rate tends to increase with wealth inequality and to decrease with how elastic savings are with respect to capital taxes. For commonly used social welfare functions, the tax on wealth is likely to be positive, as higher income households tend to hold more wealth. This happens because of precautionary savings. Optimal labor taxes are determined by the same forces in a life cycle setting, as in the static models of labor income taxation (Diamond (1998), Saez (2001)). An additional force arises, however, in the presence of capital income taxes. In particular, the formulas are adjusted by how much labor taxes influence wealth accumulation and savings decisions.

We then turn to numerical simulations of optimal policies, using estimates from the recent literature on earnings dynamics over the life cycle (Karahan and Ozkan (2012)). We can confirm the intuition from the theoretical analysis of the model that the government taxes capital income at positive rates. If capital taxes are allowed to be age-dependent, they increase over the life cycle, as wealth concentration increases. If labor taxes are allowed to be age-dependent, they also increase over the life cycle, as labor income inequality increases.

The normative theory of capital taxation is a controversial subject in the economics literature. An important and influential benchmark in the academic and popular debate on capital taxation is the simple life cycle model by Atkinson and Stiglitz (1976). In their seminal framework, individuals differ in labor abilities, preferences are (weakly) separable between consumption and leisure and there is a retirement period individuals want to save for. The main result is that a nonlinear labor income tax is the more efficient tool for redistribution and the optimal capital tax rate is zero.[2] One intuition, e.g. recently articulated in work by Piketty and Saez (2013b), for this strong normative prescription is that heterogeneity is one-dimensional in the baseline Atkinson and Stiglitz (1976) framework, making it sufficient to use only one instrument (the labor tax) applied to the source of heterogeneity directly, without distorting other margins of behavior (savings). Many empirical papers have documented that the the distribution of wages and labor income changes dramatically as we follow a cohort over their life cycle (among many Heathcote, Perri, and Violante (2010)). With changes in inequality over the life cycle, heterogeneity becomes multidimensional. We argue and show that with a realistic life cycle structure with changing inequality, a natural role for the capital tax arises on top of nonlinear labor income taxes.[3] In our model, wealth inequality arises as inequality in labor productivities changes both because of forecastable reasons

[2]This result has been generalized by Kaplow (2006) to the case where the labor income tax is not chosen optimally by the government.

[3]Some paper have emphasized that time preference heterogeneity may justify positive capital taxation in models similar to the Atkinson-Stiglitz framework; see Saez (2002), Diamond and Spinnewijn (2011) and Golosov, Troshkin, Tsyvinski, and Weinzierl (2012).

and because of idiosyncratic risk. Individuals engage in precautionary savings and a wealth distribution arises – as is well understood from the incomplete markets literature (Aiyagari (1994)). The optimal capital tax rate tends to increase with wealth inequality, as the wealthy also consume more and consequently have a lower social marginal welfare weight.

A relatively recent literature, sometimes called the *New Dynamic Public Finance* (NDPF), has explicitly taken into account how inequality evolves over the life cycle because of idiosyncratic risk, as we do in this paper. Our current work is particularly related to the papers by Golosov, Troshkin, and Tsyvinski (2011b) and Farhi and Werning (2013), who characterize history-dependent optimal labor and savings distortions in such dynamic environments. In contrast, to their work we limit our attention to simple tax structures, which are only allowed to condition on current earnings (and potentially age). We view these two approaches as clearly complementary, since the NDPF approach has the advantage that history-dependent tax systems are more powerful to raise welfare, whereas our approach has the advantage of being within the realm of current tax practices.

There is an increasing interest in tax reforms, which would move current tax policies towards conditioning on the age of the taxpayer. Weinzierl (2011) and Bastani, Blomquist, and Micheletto (2011) study optimal age-dependent labor income taxation in a discrete type model with a small number of types. They find large welfare gains from age-dependent labor income taxes and find them to be increasing with age. In contrast, we develop a first-order approach, which allows to study a much richer type space, in line with the continuous version of the Mirrlees model (Saez 2001, Golosov, Troshkin, and Tsyvinski (2011b)). Thus, we are able to optimize over a fully nonlinear labor income tax schedule and characterize it theoretically, connecting it precisely with the static literature. In addition our focus is also on age-dependent capital income taxation, which we find to increase over the life cycle.[4] Conesa, Kitao, and Krueger (2009), in tradition with the Ramsey approach to optimal taxation, study optimal labor and capital income taxes in a computational life cycle framework. While our approach shares some features from a Ramsey type of exercise, we allow labor income taxes to be an arbitrarily non-linear function in the Mirrlees tradition and theoretically highlight the forces driving labor and capital taxation.

[4]Referring to the study of Weinzierl (2011), Banks and Diamond (2011) emphasize that further research in this area 'seems to have a good probability of leading to significant policy improvements'.

3.2 The Formal Framework

3.2.1 The Model

We consider a life cycle framework with T periods where individuals at a certain point in time t are characterized by their productivity θ_t. Further, we define the history of shocks as $\theta^t = (\theta_1, \theta_2, \ldots, \theta_t)$. In each period, individuals make a savings and a labor supply decision. Flow utility is given by

$$U(c_t, y_t, \theta_t) = U\left(c_t - \Psi\left(\frac{y_t}{\theta_t}\right)\right),$$

where $U'' < 0$, $\Psi'' > 0$, c_t is consumption in period t and y is gross income in period t. $\frac{y_t}{\theta_t}$ captures labor effort. For brevity, we sometimes write $R_t = c_t - \Psi\left(\frac{y_t}{\theta_t}\right)$. Abusing notation a little bit, we will sometimes write the utility function or marginal utility as function of the history of shocks, i.e. $U(\theta^t)$.

Importantly, the functional form of U eliminates income effects on labor supply, while preserving risk-aversion. As we explain in more detail below, this assumption is crucial for the tractability of the dynamic optimal tax problem. The empirical literature using detailed micro data sets has typically not rejected a zero income elasticity or found very small effects (see Gruber and Saez (2002) for the US or a recent paper by Kleven and Schultz (2012) using the universe of danish tax records).[5] Eliminating income effects has also proven to be a key simplification in making progress on the theoretical and computational side in public finance models and especially in optimal taxation problems (Diamond (1998), Golosov, Troshkin and Tsyvinski (2011)).

We assume that agents already differ in the first period. The conditional density function (*cdf*) of the initial distribution of productivities is denoted by $F(\theta)$ and captures the ex-ante heterogeneity of agents. In the following, one should think about this heterogeneity as the the level of heterogeneity of individuals at age of roughly 25.

In the following periods, productivities evolve stochastically over time according to a Markov process. The respective *cdf* is $F(\theta_t|\theta_{t-1})$. Further let $H(\theta^t)$ be the measure over the history θ^t.[6] We consider a small open economy, so the interest on savings r is fixed. Further, we assume incomplete markets in a sense that individuals only have access to risk-free one period bonds.[7]

[5] In macroeconomics, this class of preferences has shown to be very useful in matching business cycle moments (Greenwood, Hercowitz, and Huffman (1988b), Mendoza and Yue (2012)).

[6] Sometimes we also write $h(\theta^t)$ in order to express the density of history θ^t, i.e. $f(\theta_t|\theta_{t-1})f(\theta_{t-1}|\theta_{t-2})\ldots\ldots f(\theta_1)$.

[7] We allow agents to borrow up to natural debt limit (see Aiyagari (1994)), which will never be binding since the utility function fulfills the Inada conditions.

In the absence of any taxes, the value function of an individual with history θ^t reads as

$$V_t(\theta_t, a_t(\theta^{t-1})) = \max U\left(c_t - \Psi\left(\frac{y_t}{\theta_t}\right)\right) + \beta \int_{\theta_{t+1}} V_t(\tilde{\theta}_{t+1}, a_{t+1}(\theta^t))dF(\tilde{\theta}_{t+1}|\theta_t)$$
$$\text{subject to } \; c_t + a_{t+1} = y_t + (1+r)a_t \tag{3.2.1}$$

where β is the discount factor. We are going to assume $\beta(1+r)$=1.

3.2.2 The Social Planner's Problem

We are interested in the Pareto efficient set of nonlinear labor income and linear capital income tax schedules that only condition on current income. Thus, we are not solving a dynamic Mirrlees taxation problem, where the government could condition policy instruments on all public information (typically the history of income and savings), but rather restrict the set of policy instruments in a Ramsey manner. However, our approach shares the feature of the Mirrlees approach that labor income taxes can be an arbitrarily nonlinear function of current income. One could call it a third-best Pareto problem, where third-best refers to the restriction on policy instruments. In the remainder of the paper, we will solely use the phrase Pareto optimal.

We consider two scenarios. In the first, the government can condition labor income tax schedules and capital taxes on time t, so $\mathcal{T} = \{\mathcal{T}_2(\cdot)\}_{t=1,\dots,T}$ and $\tau = \{\tau_t\}_{t=1,\dots,T}$. This is equivalent to *age-dependent* income taxation. In the second scenario, we study income tax functions, which are independent of time/age. This will be included in a future version of the paper. In both cases, we restrict optimal capital taxes to be linear.

The preferences of the social planner are described by the set of Pareto weights $\{\tilde{f}(\theta)\}_{\theta\in[\underline{\theta},\overline{\theta}]}$. The cumulative Pareto weights are defined by $\tilde{F}(\theta) = \int_{\underline{\theta}}^{\theta} \tilde{f}(\tilde{\theta})d\tilde{\theta}$. Further, the set of weights are restricted such that $\tilde{F}(\overline{\theta}) = 1$. Different sets of Pareto weights refer to different points on the Pareto frontier. The set of weights where $\tilde{f}(\theta) = f(\theta) \;\forall\; \theta$, e.g., refers to the Utilitarian planner.[8]

Before writing down the problem of the planner, we refer to a special property that any equilibrium, with taxes as defined above, has:

Lemma 3.2.1. *The optimal labor supply of an individual given taxes $\mathcal{T} = \{\mathcal{T}_2(\cdot)\}_{t=1,\dots,T}$ and $\tau = \{\tau_t\}_{t=1,\dots,T}$ will only be a function of the current shock, i.e. y_t is only a function of θ_t.*

This Lemma is a direct consequence of our preference assumption and will render the following tax problem tractable:

[8]Similar as $H(\theta^t)$ and $h(\theta^t)$, we sometimes use $\tilde{H}(\theta^t)$ and $\tilde{h}(\theta^t)$ to express the Pareto weights for individuals with certain histories.

$$\max_{\mathcal{T},\tau} \int_{\theta_1} V(\theta_1) d\tilde{F}(\theta_1) \tag{3.2.2}$$

with

$$V(\theta_1) = U\left(c_1(\theta_1) - \Psi\left(\frac{y_1(\theta_1)}{\theta_1}\right)\right) + \sum_{t=2}^{T} \beta^{t-1} \int_{\theta^t \in \mathcal{B}_2(\theta_1)} U\left(c_t(\theta^t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) dH(\theta^t), \tag{3.2.3}$$

where $\mathcal{B}_2(\theta_1)$ is the set of all θ^t that contain θ_1 as its first element. The government has to balance the budget intertemporally

$$\sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\theta^{t-1}} \int_{\theta_t} \left(\mathcal{T}_2(y_t(\theta_t)) + \tau_t^k(1+r)a_t(\theta^{t-1})\right) dF_t(\theta_t|\theta^{t-1}) dH(\theta^{t-1}) = \mathcal{R} \tag{3.2.4}$$

where $\mathcal{R}$ is some exogenous revenue requirement of the government.

And the government has to take into account individual behavior, i.e. $\{c_t(\theta^t), y_t(\theta_t)\}$ solve

$$V_t(\theta_t, a_t(\theta^{t-1})) = \arg\max_{c_t, y_t} \; U\left(c_t - \Psi\left(\frac{y_t}{\theta_t}\right)\right) + E_t\left[V_{t+1}(\theta^t, a_{t+1}(\theta^t))\right]$$
$$\text{where: } c_t + a_{t+1}(\theta^t) = y_t - \mathcal{T}_2(y_t) + (1+r)(1-\tau_t^k)a_t(\theta^{t-1}) \tag{3.2.5}$$

Constraint (3.2.5) makes the solution of the problem with Lagrangian methods nontrivial. In the following subsection, we argue that (3.2.5) can be replaced by a set of first-order conditions for a_t and y_t. a_t is net-wealth of an individual at the beginning of period t, and we will refer to it throughout as wealth, savings or capital, used freely and interchangeably.

3.2.3 First-Order Approach

The set of first-order conditions for the individual problem (3.2.5) are standard. For the labor supply decision, we have $\forall t$ and $\forall \theta_t$:

$$1 - \mathcal{T}'(y_t(\theta_t)) = \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\frac{1}{\theta_t}. \tag{3.2.6}$$

For the intertemporal consumption decision we have $\forall t = 1, ..., T-1$ and $\forall \theta^t$:

$$U'\left(c_t(\theta^t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) = \int_{\theta_{t+1}} U'\left(c_t(\theta^t, \theta_{t+1}) - \Psi\left(\frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right)\right) dF_{t+1}(\theta_{t+1}|\theta_t). \tag{3.2.7}$$

with

$$c_t(\theta^t) = y_t(\theta_t) - \mathcal{T}_2(y_t(\theta_t)) + (1+r)(1-\tau_t^k)a_t(\theta^{t-1}) - a_{t+1}(\theta^t)$$

Needless to say, these conditions are only necessary and not sufficient for the agents choices to be optimal. Due to assumption about preferences, the second order conditions are of particularly simple form. The derivative of the first-order condition of labor supply with respect to consumption, i.e. the cross derivative of the value function, is zero. By symmetry of the Hessian, the same holds for the derivative of the Euler equation with respect to labor supply. Thus, the minor diagonal of the Hessian matrix contains only zeros. For (3.2.6) and (3.2.7) to represent a maximum, only the second derivatives of the value function with respect to labor supply and consumption have to be ≤ 0. For labor supply, a familiar argument from the standard Mirrlees model implies that this holds if and only if

$$y'(\theta_t) \geq 0. \tag{3.2.8}$$

The second order condition for savings is always fulfilled due to concavity of the utility function. Hence, (3.2.6) and (3.2.7) represent a maximum whenever $y'(\theta_t) \geq 0$. As $y'(\theta_t) \geq 0$ even implies global concavity, (3.2.6) and (3.2.7) even represent a global maximum if $y'(\theta_t) \geq 0$ holds.

Lemma 3.2.2. *When choosing $\mathcal{T} = \{\mathcal{T}_2(\cdot)\}_{t=1,\ldots,T}$ and $\tau = \{\tau_t\}_{t=1,\ldots,T}$ to maximize (3.2.2) subject to (3.2.4) and (3.2.5), the last constraint can be replaced by (3.2.6), (3.2.7) and (3.2.8).*

Incorporating (3.2.6) into a Lagrangian, however, is still problematic as it contains $\mathcal{T}'$, i.e. the derivative of the function with respect to which we want to maximize. To tackle this problem, we make use of the following derivative

$$\frac{\partial\left(y_t(\theta_t) - \mathcal{T}_t(y_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial\theta_t} = y_t'(\theta_t)(1-\mathcal{T}_2'(y_t)) - \Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right)\left[\frac{y_t'(\theta_t)}{\theta_t} - \frac{y_t(\theta_t)}{\theta_t^2}\right].$$

Inserting (3.2.6) into this derivative yields:

$$\frac{\partial\left(y_t(\theta_t) - \mathcal{T}_2(y_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial\theta_t} = \Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right)\frac{y_t(\theta_t)}{\theta_t^2}. \tag{3.2.9}$$

Thus, (3.2.9) implies (3.2.6). As we show in the appendix, (3.2.9) can easily be incorporated into the Lagrangian. Further, when solving for optimal policies, we will not incorporate the monotonicity constraint (3.2.8) into the Lagrangian as is standard practice in the optimal tax literature. In the numerical simulations we will ex-post check whether the condition is fulfilled or not. The Lagrangian and all first-order conditions for the age-dependent and age-independent case are stated in the appendix.

3.3 Pareto Optimal Taxes

We start by deriving and discussing optimal taxes in a two period model, so $T = 2$. The reason is that optimal taxes in the two periods case are much easier to derive and, without loosing too much economic intuition in the interpretation in comparison to the T period case. Building on our results we then briefly discuss formulas for the general case.

3.3.1 The Two Period Model

Labor Taxes

We derive optimal tax formulas for our *dynamic problem* following an intuitive tax reform approach as in the *static* Mirrlees literature, pioneered by Piketty (1997) and further developed by Saez (2001). In the appendix we document how to derive the tax formulas from first-order conditions of the optimal control problem we spelled out in the last section.

We start with the age-dependent tax in period two. Consider an infinitesimal increase in the marginal tax rates $d\mathcal{T}_2'$ within an income interval of infinitesimal length dy_2 around some income level y_2. First, such a change in marginal tax rates triggers a *mechanical* increase in tax revenue by raising the tax obligation for all individuals at age 2 with higher income in that period. The overall effect on the government budget is:

$$d\mathcal{T}_2' dy_2 \int_{\theta_1} (1 - F(\theta_2|\theta_1)) dF(\theta_1). \tag{3.3.1}$$

People affected by this marginal tax rate increase loose; this is taken into account by the government:

$$d\mathcal{T}_2' dy_2 \int_{\theta_1} \int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF(\tilde{\theta}_2|\theta_1) d\tilde{F}(\theta_1), \tag{3.3.2}$$

where $\tilde{F}(\theta_1)$ the distribution of Pareto enters, as different social welfare functions will place a different value on redistribution. The tax increase will also trigger behavioral responses. First, on the labor supply margin, which is captured by:

$$\mathcal{T}_2' \frac{\partial y_2}{\partial(1-\mathcal{T}_2')} d\mathcal{T}_2' \times Mass(y_2). \tag{3.3.3}$$

Elastic labor supply tends to decrease optimal tax rates.[9] In addition in a dynamic model, savings behavior will change:

$$\tau^k(1+r)\int_{\theta_1} \frac{\partial a_2(\theta_1)}{\partial \mathcal{T}_2'(y)} dF(\theta_1) \tag{3.3.4}$$

An increase in taxes at the old age will induce the young to save more. If capital taxes τ^k are different from zero, this has a first-order effect on the government's budget.

Using a tax perturbation approach also clarifies how the absence of income-effects simplifies the problem, as a the tax increase does not trigger any direct labor supply responses in period one, since the labor supply functions just depend on the current ability level and tax rates in the current period.

At an optimum, weighing (3.3.4), (3.3.3) and (3.3.1) by the (present-value) of the marginal value of public funds denoted by λ and adding all terms, one gets the following optimal tax formula.

Proposition 3.3.1. *Pareto optimal marginal labor income tax rates for the old satisfy::*

$$\frac{\mathcal{T}_2'(y_2(\theta_2))}{1-\mathcal{T}_2'(y_2(\theta_2))} = \left(1+\frac{1}{\varepsilon(\theta_2)}\right)\frac{1}{\theta_2\int_{\theta_2} f(\theta_2|\theta_1)dF(\theta_1)} \times \left[\mathcal{M}_2(\theta_2)+\mathcal{S}_2(\theta_2)\right],$$

where the total mechanical effect equals

$$\mathcal{M}_2(\theta_2) = \int_{\theta_1}\int_{\theta_2}^{\bar{\theta}} \left(1-\frac{U`(\theta_1,\tilde{\theta}_2)}{\lambda}\frac{\tilde{f}(\theta_1)}{f(\theta_1)}\right) dF(\tilde{\theta}_2|\theta_1)d\tilde{F}(\theta_1)$$

and the savings response equals

$$\mathcal{S}_2(\theta_2) = \frac{\tau^k}{(1+r)}\int_{\theta_1} \frac{\partial a_2(\tilde{\theta}_1)}{\partial \mathcal{T}_2'(y_2(\theta_2))} dF(\theta_1).$$

The optimal tax formula follows the same logic as in the static literature. i.e. it is decreasing in the elasticity of labor supply and trades-off how an additional dollar in the hand of the government is weighted against the welfare loss of individuals who face

[9] It can be show that the total mass of individuals whose tax income is affected can be written as: $\frac{\theta_2}{y_2(1+\varepsilon)}\int_{\theta_1} f(\theta_2|\tilde{\theta}_1)dF(\theta_1)dy_2$

higher taxes. The novel behavioral response in a dynamic model is the savings margin $\mathcal{S}_2(\theta_2)$. Notice that $\mathcal{S}_2(\theta_2)$ has the same sign as τ^k, because $\frac{\partial a_2(\tilde{\theta}_1)}{\partial \mathcal{T}_2'(y_2(\theta_2))} > 0$. We argue in the next section and confirm in our numerical simulations that a positive $\tau^k > 0$ is likely with a realistic life cycle environment, so that $\mathcal{S}_2(\theta_2)$ is likely to be positive as well. In contrast, the savings response w.r.t. to a tax increase at some skill level θ_1 for the young is given by:

$$\mathcal{S}_1(\theta_1) = \frac{\tau^k}{(1+r)} \int_{\theta_1}^{\overline{\theta}} \frac{\partial a_2(\tilde{\theta}_1)}{\partial \mathcal{T}_1'(y_1(\theta_1))} dF(\tilde{\theta}_1).$$

Because of $\frac{\partial a_2(\tilde{\theta}_1)}{\partial \mathcal{T}_1'(y_1(\tilde{\theta}_1))} < 0$, $\mathcal{S}_1(\theta_1)$ is negative with a positive capital tax. Taken together, the presence of positive capital taxation will tend to increase labor taxes over the life cycle.

If labor taxes are restricted to be independent of age, an almost identical formula as in Proposition 3.3.1 applies. Intuitively, the underlying trade-offs are same for the government, with the mechanical and labor supply effects being weighted over both periods. The effect of labor taxes on savings will in general be ambiguous – as our discussion on age-dependent labor taxes highlights, higher taxes can reduce savings by an income effect when individuals are young but might also increase savings as individuals anticipate higher taxes later in life. Finally notice that if the government is not able or not willing to tax capital so $\tau^k = 0$, the savings responses $\mathcal{S}_i$ are also zero. In this case, the optimal tax formulas collapse to their static counterparts (Diamond 1998), adjusted by weighing mechanical and behavioral responses across both periods.

Capital Taxes

In the two period model there is no difference between age-dependent and age-independent capital taxation, as young agents start with zero wealth and capital is only taxed in period two. We will again use a perturbation argument, looking at a small change in the capital tax rate $d\tau^k$.

This will increase government's revenue by

$$\frac{1}{(1+r)} \int_{\theta_1} a_2(\tilde{\theta}_1) dF(\tilde{\theta}_1).$$

and decrease utility of individuals, which is valued by

$$\int_{\theta_1} a_2(\tilde{\theta}_1) \int_{\theta_2}^{\overline{\theta}} U'(\tilde{\theta}_1, \tilde{\theta}_2) dF(\tilde{\theta}_2 | \theta_1) d\tilde{F}(\tilde{\theta}_1).$$

It will also discourage savings and thereby savings tax revenue, given by:

$$\frac{\tau^k}{(1+r)}\int_{\theta^1}\frac{\partial a_2(\tilde{\theta}_1)}{\partial \tau^k}dF(\tilde{\theta}_1)$$

The absence of income effects implies that labor supply will not change in response to the small tax increase. Adding and collecting terms again and weighing the revenue effects by the marginal value of public funds λ yields the formula for the optimal capital tax rate:

Proposition 3.3.2.

$$\frac{\tau^k}{1-\tau^k}=\frac{(1+r)\int_{\theta_1}a_2(\tilde{\theta}_1)\left[f_1(\tilde{\theta}_1)-\int_{\theta_2}\frac{U'(\tilde{\theta}_1,\tilde{\theta}_2)\tilde{f}_1(\tilde{\theta}_1)}{\lambda}f(\tilde{\theta}_2|\theta_1)\right]d\tilde{\theta}_2 d\tilde{\theta}_1}{\int_{\theta_1}a_2(\tilde{\theta}_1)\zeta(\tilde{\theta}_1)dF_1(\tilde{\theta}_1)},$$

where $\zeta(\tilde{\theta}_1)$ is the elasticity of savings w.r.t the net of tax rate $1-\tau^k$.

The optimal taxation of capital follows a very simple and intuitive equity-efficiency trade-off, as is standard in the public finance literature. It is decreasing in the weighted elasticity of savings w.r.t to the net of tax rate $1-\tau^k$. The tax rate is higher, the more the government values redistribution from savers to non-savers. Note that is motive is independent of the presence of idiosyncratic uncertainty. To see this, suppose that in the second-period there would be no labor supply so individuals would effectively retire. To smooth consumption, higher income individuals would save more. For standard social welfare weights, the government would tax savings at a positive rate. In future work, we plan to understand how our optimal tax formula nests the zero capital tax result from Atkinson and Stiglitz (1976), which one obtains in the presence of separable preferences, a retirement period and no idiosyncratic risk.

In the present life cycle model with idiosyncratic risk, wealth inequality arises not because of a retirement period, but because individuals engage in precautionary savings, as is well-understood from the work of Bewley, Aiyagari and many other following their footsteps. As wealth will typically be correlated with a low marginal social welfare weight, Proposition 3.3.2 clarifies while a role for positive capital taxation is likely to arise in a life cycle model with changing inequality.

In a recent paper, Jacobs and Schindler (2012) show that in a two period model with linear labor taxes, a similar role for the capital tax may arise. In their framework, a positive capital tax insures in the presence of idiosyncratic risk. In their framework, capital taxes also have the positive effect of boosting labor supply in the second period.

3.3.2 The T Period Case

We next present the optimal tax formulas in the general T period case.

Labor Taxes

We again present the formula for a typical marginal tax rate on labor income, if age-dependent taxes are available for the government.

Proposition 3.3.3. *Age-dependent Pareto optimal marginal labor income tax rates satisfy*

$$\frac{\mathcal{T}_t'(y_t(\theta_t))}{1-\mathcal{T}_t'(y_t(\theta_t))} = \left(1+\frac{1}{\varepsilon(\theta_t)}\right)\frac{1}{\theta_t \int_{\theta^{t-1}} f(\theta_t|\theta_{t-1})dH(\theta^{t-1})} \times [\mathcal{M}_t(\theta_t)+\mathcal{S}_t(\theta_t)]$$

where the total mechanical effect equals

$$\mathcal{M}_t(\theta_t) = \int_{\theta^{t-1}}\int_{\theta_t}^{\overline{\theta}} dF(\tilde{\theta}_t|\theta_{t-1})dH(\theta^{t-1}) - \lambda \int_{\theta^{t-1}}\int_{\theta_t}^{\overline{\theta}} U'(R_t(\theta^{t-1},\tilde{\theta}_t))dF(\tilde{\theta}_t|\theta_{t-1})d\tilde{H}(\theta^{t-1})$$

and the savings response equals

$$\begin{aligned}\mathcal{S}_t(\theta_t) = &-\int_{\theta^{t-1}}\int_{\theta_t}^{\overline{\theta}} \mu_t(\theta^{t-1},\theta_t)U''(R_t(\theta^{t-1},\theta_t))d\theta_t d\theta^{t-1}\\ &+(1-\tau_t^k)\int_{\theta^{t-1}}\mu_{t-1}(\theta^{t-1})\int_{\theta_t}^{\overline{\theta}} U''(R_t(\theta^{t-1},\theta_t))dF_t(\theta_t|\theta^{t-1})d\theta^{t-1},\end{aligned}$$

where $\mu_t(\theta^t)$ is the multiplier on the Euler equation of an agent with history θ^t.

The main insight is that the same basic forces underlying optimal taxes in the two period model also determine tax rates in the T period case. Higher mechanical revenue effects tend to increase tax rates, the labor supply elasticity tends to decrease tax rates. Labor taxes will also influence savings behavior, as captured by $\mathcal{S}_t(\theta_t)$. In the first period t=1 this will lead to lower tax rates whereas in the last period t=T it will lead to higher tax rates. If capital taxation is unavailable to the government, it again follows that $\mathcal{S}_t(\theta_t) = 0$ in all periods. For a future version of this paper, we are planning to express the savings responses $\mathcal{S}_t(\theta_t)$ as a direct function of the behavioral responses, so as a function of $\frac{\partial a_i}{\partial \mathcal{T}_j'(y_j(\theta_j))}$ for all $i,j = 1,\dots,T$. In the age-independent case, an almost identical formula applies.

Capital Taxes

A small increase in the age-dependent capital tax at age t will again trigger mechanical revenue effects for the government and reduce the welfare of wealth holder at age t. As in the case of the age-dependent labor tax, the behavioral response on savings will not only be limited to wealth holdings a_t, but also have effects on the whole sequence of savings throughout the life cycle $a_2, \dots, a_T$. The following formula can be derived:

Proposition 3.3.4.

$$
\begin{aligned}
(1-\tau_t^k) = {} & \frac{1}{\beta(1+r)^2\mu_{t-1}(\theta^{t-1})\int_{\theta_t} U''(R_t(\theta^{t-1},\theta_t))a_t(\theta^{t-1})dF_t(\theta_t|\theta_{t-1})d\theta^{t-1}} \\
& \Bigg[-\frac{\lambda}{(1+r)^{t-2}}\int_{\theta^{t-1}} a_t(\theta^{t-1})dH(\theta^{t-1}) \\
& + \beta^{t-1}(1+r)\int_{\theta^{t-1}} a_t(\theta^{t-1})\int_{\theta_t} U'(R_t(\theta^{t-1},\theta_t))dF_t(\theta_t|\theta_{t-1})d\tilde{H}(\theta^{t-1}) \\
& + (1+r)\int_{\theta^{t-1}} a_t(\theta^{t-1})\int_{\theta_t} \mu_t(\theta^{t-1},\theta_t)U''(R_t(\theta^{t-1},\theta_t))d\theta_t d\theta^{t-1} \\
& - \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1})\int_{\theta_t} U'(R_t(\theta^{t-1},\theta_t))dF_t(\theta_t|\theta_{t-1})d\theta^{t-1} \Bigg],
\end{aligned}
$$

where $\mu_t(\theta^t)$ is the multiplier on the Euler equation of an agent with history θ^t.

In comparison to the two period case, with age-dependent capital taxation adjustments on the history of Euler equations should now include spillovers on capital tax revenue in all periods. For a future version of this paper we plan to express these adjustments on the individuals' Euler equation as savings responses.

3.4 A Numerical Exploration

We now numerically simulate optimal policies for a T=3 period economy.

3.4.1 Inequality over the Life Cycle and Parameters

There is large literature on the estimation of earnings dynamics over the life cycle – see Meghir and Pistaferri (2011) and Jappelli and Pistaferri (2010) for recent surveys. For the parameterization of our model, we use the recent empirical approach taken by Karahan and Ozkan (2012). In their analysis, which we describe in more detail in the appendix, they estimate the persistence of permanent shocks as well as the variance of permanent and transitory income shock for US workers. Innovatively and in contrast to most pervious work in this strand of literature, they allow these parameter to be age-dependent and changing over the life cycle.

They find two structural breaks in how the key parameters change over the life cycle, giving three age groups, in which income dynamics are governed by the same risk

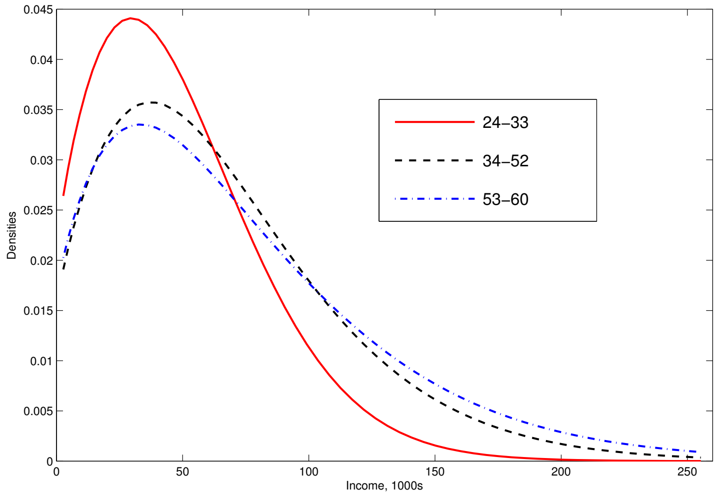


Figure 3.1: Income Distribution for Each Age Group

parameters. The age groups are 24-33, 34-52 and 53-60 years old respectively. Given the estimates of Karahan and Ozkan (2012), we simulate millions of labor income histories. The first period is interpreted as the time between 24-33, the second as 34-53 and the last one from 53-60 years. After having simulated those earnings histories, we calibrate the skill distributions Figure 3.1 shows the three cross-sectional income distribution for each age group from our simulated data. It becomes clear how inequality evolves over the life cycle. In the older age groups there are more people with top incomes, but also more with low labor earnings than in the young age group. Figure 3.2 shows three conditional income distribution for the middle age-group, conditioning on earning \$14,000, \$30,000 and \$180,000 respectively. The role of both persistence and risk for earnings becomes clear from this picture. As a last step to complete the calibration of the model, we calibrate all conditional skill distributions from their income counterparts, as suggested by Saez (2001).

We assume the utility function to be of the form

$$U = \frac{\left(C - \frac{L^{1+\varepsilon}}{1+\varepsilon}\right)^{1-\gamma}}{1-\gamma}. \tag{3.4.1}$$

For the benchmark, we set $\varepsilon = 3$, implying a Hicks elasticity of 0.33 (Chetty (2012a)) and set $\gamma = 1.5$ (Chetty (2006)). The yearly interest rate is 4% and $\beta = \frac{1}{1+r}$. Since the three periods have different lengths, we adjust for it when picking discount factor and interest rates for our three period simulations.

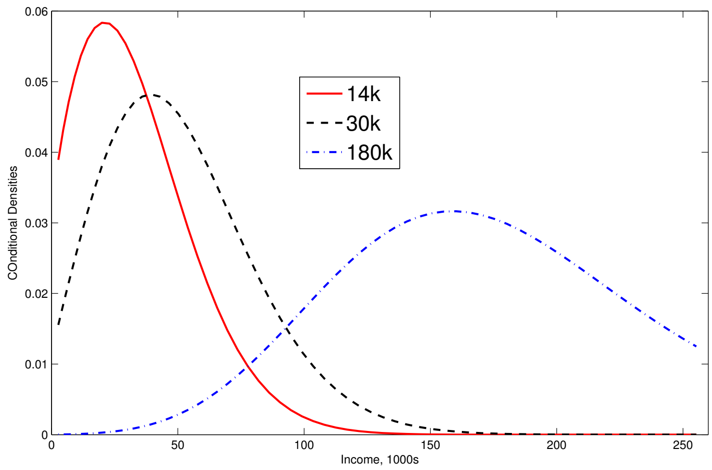


Figure 3.2: Conditional Income Distributions, Middle Aged Workers

3.4.2 Results in the Benchmark case

We present results for a utilitarian social welfare function. We calculate optimal policies for four cases: age-dependent taxes available or not and wealth/capital income taxes available or not.

In Figure 3.3 we illustrate optimal marginal labor income tax rates, both age-dependent and age-independent, for the case when wealth taxation is available to the government. First, all marginal tax rates are decreasing over the income distribution, reflecting that the income distributions have a log-normal shape. This marginal tax rate regressivity is well-understood from the static literature on optimal income taxation (Diamond (1998)).[10] Second, labor income taxes are increasing over the life cycle. The intuition is that labor income inequality is much higher at later points in the life cycle. This leads to bigger mechanical revenue effects for the government when raising tax rates. Third, intuitively in the case in which the government is restricted to set age-independent taxes, optimal marginal tax rates are roughly an average of the age-dependent taxes. Fourth, in the case in which wealth taxes are not available as an instrument, the resulting labor tax rates look very similar to the ones in Figure 3.3. So there is only a moderate effect of the behavioral response of savings w.r.t. to labor taxes.

For optimal capital tax rates, we present our results as the yearly optimal tax rates on capital income, so that revenue raised from an individual with wealth a is given by $\tau_i^k Ra$, where R is the yearly rate-of-return. We find significant and positive optimal tax rates on capital income, confirming the theoretical intuition of using the instrument

[10] Notice that our income distributions have no Pareto tails. From figure 3.1 it becomes obvious that tails tend to become much thicker over the life cycle, but never thick enough to resemble a Pareto tail as in Saez (2001).

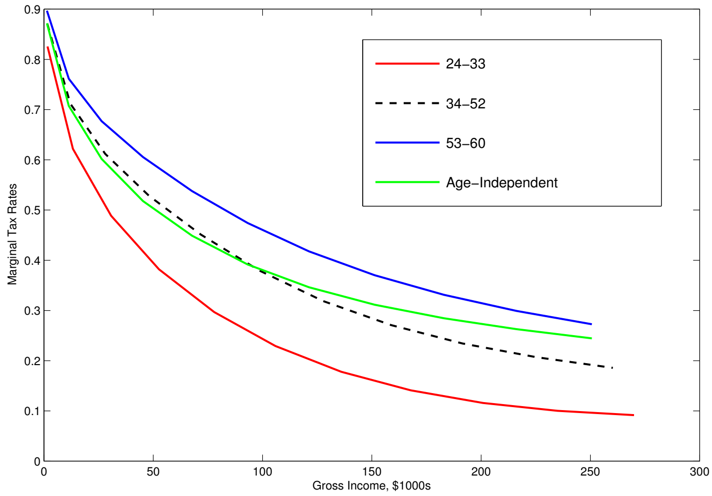
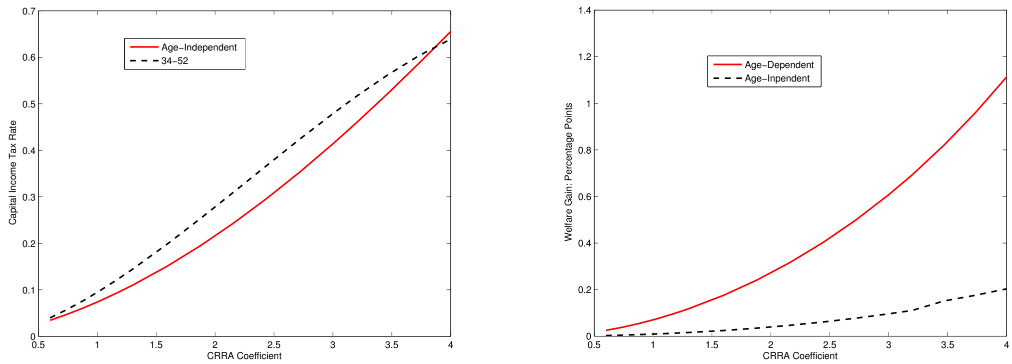


Figure 3.3: Optimal Marginal Income Tax Rates: Utilitarian Case

as a redisitributive tool. In the age-independent case, the resulting optimal tax rate is $\tau_i^k = 13.70\%$. If allowed to be age-dependent, the optimal yearly income tax rates are $\tau_M^k = 18.05\%$ and $\tau_O^k = 28.17\%$. Capital income taxes are increasing in age, as wealth inequality increases over the life cycle, increasing the value for the government to tax wealth. The age-independent tax is not an average its age-dependent counterparts in this case. The reason is that individuals' savings behavior is different in the case with age-dependent taxes. The mean and the variance of the wealth distribution tend to be higher with age-dependent labor taxes, as individuals save anticipating higher labor tax rates. The higher wealth inequality increases the effectiveness of capital income taxes in the age-dependent case even further.

3.4.3 Welfare Gains and Comparative Statics: Risk Aversion and Intertemporal Elasticity of Substitution

We next test the sensitivity of optimal policies to γ, which controls both risk-aversion and the intertemporal elasticity of substitution (IES). Figure 3.4(a) illustrates that the value of the capital income tax rate is quite sensitive to the value of γ. The optimal age-independent rate gets as high as 60% for values of γ around four. The same is true for age-dependent capital tax rates. Higher Risk-aversion tends to increase capital taxes, as it increases the value placed on redistribution and insurance. Simultaenously, the IES decreases, decreasing the elasticity of savings with respect to the tax rate. Both effects explain the upward sloping profile.



(a) Risk-Aversion and Capital Taxes (b) Welfare Gains of Capital Taxation

Figure 3.4: Wealth Taxes

Figure 3.4(b) show the consumption equivalent welfare gains in percentage points of being able to tax wealth. These gains are also increasing in risk-aversion and are higher in the age-dependent case, underscoring the complementarity of being able to condition both labor and wealth taxes on age. The gains range from 0.025% to 1.113% points as γ goes from 0.5 to 4 in the age-dependent case and from 0.003% to 0.203% points in the age-independent case. For comparison the gains from being able to condition on annual tax on age lie between 0.200% to 1.310% when wealth taxation is available and 0.170% to 0.350% when it is not.

3.4.4 Different Social Welfare Function and the Role of Insurance

In a framework with heterogeneous agents, there is no correct or wrong normative objective. Whereas the Utilitarian case is a popular and important benchmark, some people might oppose this criterion as being to redistributive in favor of young low income individuals. We now look at a case where Pareto weights are such that the planner favors redistribution from poor to rich agents. This case is illustrated in Figure 3.5: the Pareto weights are shifted to the right as compared to the density, i.e. social marginal welfare weights are increasing in ability. In this case (and $\gamma = 1.5$) , first-period marginal labor income tax rates negative for lower incomes and low and flat higher incomes as can be seen in Figure 3.6. In contrast, they are strictly positive in the subsequent periods. This illustrates the role of taxation as an insurance device. When individuals are young, high-income individuals like low labor taxes. In the presence of idiosyncratic labor market risk, some of those initially well-off draw negative labor income shocks, moving them down in the distribution. The government takes that into account when designing tax schedules for older ages.

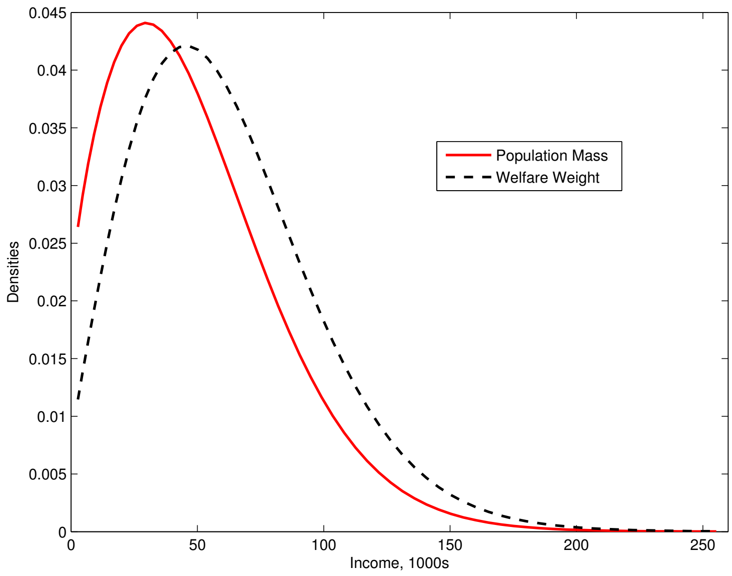


Figure 3.5: Welfare Function More in Favor of High-Income Young

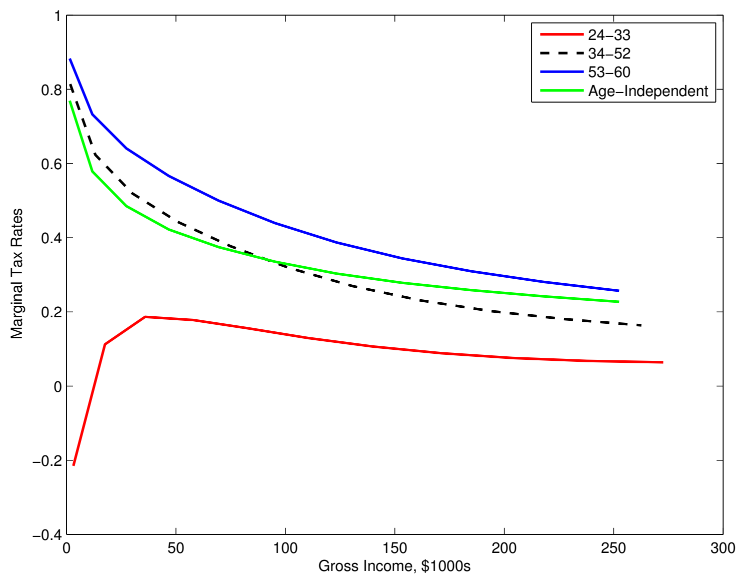


Figure 3.6: Optimal Labor Income Taxes

The same logic applies for optimal capital income taxes: in the middle period it is subsidized to redistribute from ex-ante low to ex-ante high types. The subsidy is $\tau_M^k = -4.55\%$. In contrast, in the last stage of the working life the tax rate is $\tau_O^k = 25.44\%$, almost as high as in the utilitarian benchmark. For comparison it is 7.45% with age-independency.

3.5 Conclusion

We have developed a formal framework to study Pareto optimal nonlinear taxation of annual labor income as well as linear taxation of capital in a framework with heterogenous agents whose skills evolve stochastically over time. This method can be used to study age-dependent and age-independent taxes. By focusing on preferences without income effects on labor supply, we developed a first-order approach to make this problem tractable also for a continuous type space.

In this dynamic environment where inequality evolves over the life cycle, we derive a novel and simple formula for optimal capital income taxes. It follows a standard public finance intuition, trading redistributive benefits versus efficiency costs resulting from behavioral responses. In our realistically calibrated numerical simulations capital income taxation plays an important role as a redistribution device in this setting. Both, savings and labor income taxes tend to increase over the lifecycle, as inequality in both dimensions increases.

In future work, we plan to disentangle the role of predictable trends in inequality versus the role of idiosyncratic labor market risk. In particular, the latter tends to increase capital income tax rates. Predictable trends in contrast, driven by heterogenous growth rates in wages across individuals can also influence optimal capital taxes, although it is not clear in which direction. In our current optimal policy simulations both are present: labor market risk and permanent differences, making it impossible to separate both effects.

In future work, it would also be valuable to disentangle the IES from risk aversion and their influence on capital taxation in a realistic life cycle setting with simple instruments as in the current paper. This could be achieved by employing a recursive preference structure – see Farhi and Werning (2008) for a first important step in that direction.

3.6 Appendix – Chapter Three

3.6.1 The Lagrangian- Age Dependent Case

$$
\begin{aligned}
\mathcal{L} = & \int_{\theta_1} \left(\sum_{t=1}^{T} \beta^{t-1} \int_{\theta^t} U \left(M_t(\theta_t) - a_{t+1}(\theta^t) + (1-\tau_t)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) dH(\theta^t) \right) d\tilde{F}_1(\theta_1) \\
& + \lambda \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\theta^{t-1}} \int_{\theta_t} y_t(\theta_t) - M(y_t(\theta_t)) + \tau_t(1+r)a_t(\theta^{t-1}) dF_t(\theta_t|\theta_{t-1}) dH(\theta^{t-1}) \\
& + \sum_{t=1}^{T-1} \int_{\theta^t} \mu_t(\theta^t) \Bigg[U'\left(M_t(\theta_t) - a_{t+1}(\theta^t) + (1-\tau_t)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) \\
& - \beta(1+r)(1-\tau_{t+1}) \int_{\theta_{t+1}} U' \Bigg(M_{t+1}(\theta_{t+1}) - a_{t+2}(\theta^t, \theta_{t+1}) \\
& + (1-\tau_{t+1})(1+r)a_{t+1}(\theta^t) - \Psi\left(\frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right)\Bigg) dF_{t+1}(\theta_{t+1}|\theta^t) \Bigg] d\theta^t \\
& + \sum_{t=1}^{T} \int_{\theta_t} \eta_t(\theta_t) \frac{\partial \left(M_t(\theta_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial \theta_t} d\theta_t - \sum_{t=1}^{T} \int_{\theta_t} \eta_t(\theta_t) \Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right) \frac{y_t(\theta_t)}{\theta_t^2} d\theta_t
\end{aligned}
$$

Partially integrating $\sum_{t=1}^{T} \int_{\theta_t} \eta_t(\theta_t) \frac{\partial \left(M_t(\theta_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial \theta_t} d\theta_t$ yields

$$
\eta_t(\overline{\theta}) \left(M_t(\overline{\theta}) - \Psi\left(\frac{y_t(\overline{\theta})}{\overline{\theta}}\right)\right) - \eta_t(\underline{\theta}) \left(M_t(\underline{\theta}) - \Psi\left(\frac{y_t(\underline{\theta})}{\underline{\theta}}\right)\right) - \int_{\theta_t} \eta_t'(\theta_t) \left(M_t(\theta_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) d\theta_t
$$

The derivatives with respect to the endpoint conditions yield $\forall t: \ \eta_t(\overline{\theta}) = \eta_t(\underline{\theta}) = 0$.

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial M_s(\theta_s)} = & - \frac{\lambda}{(1+r)^{s-1}} \int_{\theta^{s-1}} f_s(\theta_s|\theta_{s-1}) dH(\theta^{s-1}) \\
& + \beta^{s-1} \int_{\theta^{s-1}} U'(R_s(\theta^{s-1}, \theta_s)) f_s(\theta_s|\theta_{s-1}) d\tilde{H}(\theta^{s-1}) \\
& + \int_{\theta^{s-1}} \mu_s(\theta^{s-1}, \theta_s) U''(R_s(\theta^{s-1}, \theta_s)) d\theta^{s-1} \\
& - \beta(1+r)(1-\tau_s) \int_{\theta^{s-1}} \mu_{s-1}(\theta^{s-1}) U''(R_s(\theta^{s-1}, \theta_s)) f_s(\theta_s|\theta^{s-1}) d\theta^{s-1} \\
& - \eta_s'(\theta_s) = 0
\end{aligned} \tag{3.6.1}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial y_s(\theta_s)} =& \frac{\lambda}{(1+r)^{s-1}} \int_{\theta^{s-1}} f_s(\theta_s|\theta_{s-1}) dH(\theta^{s-1}) \\
& - \beta^{s-1} \int_{\theta^{s-1}} U'(R_s(\theta^{s-1}, \theta_s)) \Psi' \left(\frac{y_s(\theta_s)}{\theta_s} \right) \frac{1}{\theta_s} f_s(\theta_s|\theta_{s-1}) d\tilde{H}(\theta^{s-1}) \\
& + \int_{\theta^{s-1}} \mu_s(\theta^{s-1}, \theta_s) U''(R_s(\theta^{s-1}, \theta_s)) \Psi' \left(\frac{y_s(\theta_s)}{\theta_s} \right) \frac{1}{\theta_s} d\theta^{s-1} \\
& - \beta(1+r)(1-\tau_s) \int_{\theta^{s-1}} \mu_{s-1}(\theta^{s-1}) U''(R_s(\theta^{s-1}, \theta_s)) \Psi' \left(\frac{y_s(\theta_s)}{\theta_s} \right) \frac{1}{\theta_s} f_s(\theta_s|\theta^{s-1}) d\theta^{s-1} \\
& - \eta_s'(\theta_s) \Psi' \left(\frac{y_s(\theta_s)}{\theta_s} \right) \frac{1}{\theta_s} - \eta_s(\theta_s) \left(\Psi' \left(\frac{y_s(\theta_s)}{\theta_s} \right) \frac{1}{\theta_s^2} + \Psi'' \left(\frac{y_s(\theta_s)}{\theta_s} \right) \frac{y_s(\theta_s)}{\theta_s^2} \right) = 0
\end{aligned} \tag{3.6.2}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_{s+1}(\theta^s)} =& \frac{\lambda}{(1+r)^{s-1}} \tau_{s+1} h(\theta^{s-1}) - \mu_s(\theta^s) U''\left(R_s(\theta^s)\right) \\
& - (1-\tau_{s+1})^2 \beta (1+r)^2 \mu_s(\theta^s) \int_{\theta_{s+1}} U''(R_s(\theta^s, \theta_{s+1})) dF_{s+1}(\theta_{s+1}|\theta^s) \\
& + (1-\tau_s) \beta (1+r) \mu_{s-1}(\theta^{s-1}) U''\left(R_s(\theta^s)\right) f_s(\theta_s|\theta^{s-1}) \\
& + (1-\tau_{s+1})(1+r) \int_{\theta_{s+1}} \mu_{s+1}(\theta^s, \theta_{s+1}) U''(R_{s+1}(\theta^s, \theta_{s+1}) d\theta_{s+1} = 0
\end{aligned} \tag{3.6.3}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \tau_s} =& \frac{\lambda}{(1+r)^{s-2}} \int_{\theta^{s-1}} a_t(\theta^{s-1}) dH(\theta^{s-1}) \\
& - \beta^{s-1}(1+r) \int_{\theta^{s-1}} a_s(\theta^{s-1}) \int_{\theta_s} U'(R_s(\theta^{s-1}, \theta_s)) dF_s(\theta_s|\theta_{s-1}) d\tilde{H}(\theta^{s-1}) \\
& - (1+r) \int_{\theta^{s-1}} a_s(\theta^{s-1}) \int_{\theta_s} \mu_s(\theta^{s-1}, \theta_s) U''(R_s(\theta^{s-1}, \theta_s)) d\theta_s d\theta^{s-1} \\
& + \beta(1+r)^2(1-\tau_s) \int_{\theta^{s-1}} \mu_{s-1}(\theta^{s-1}) \int_{\theta_s} U''(R_s(\theta^{s-1}, \theta_s)) a_s(\theta^{s-1}) dF_s(\theta_s|\theta_{s-1}) d\theta^{s-1} \\
& + \beta(1+r) \int_{\theta^{s-1}} \mu_{s-1}(\theta^{s-1}) \int_{\theta_s} U'(R_s(\theta^{s-1}, \theta_s)) dF_s(\theta_s|\theta_{s-1}) d\theta^{s-1} = 0
\end{aligned} \tag{3.6.4}
$$

3.6.2 Labor Tax Rates

Dividing (3.6.2) by $\Psi' \frac{1}{\theta_s}$ and adding (3.6.1) yields

$$
\frac{\mathcal{T}_s'(y_s(\theta_s))}{1-\mathcal{T}_s'(y_s(\theta_s))} = \frac{\varepsilon(\theta_s)+1}{\varepsilon(\theta_s)} \frac{\eta_s(\theta^s)}{\lambda \frac{1}{(1+r)^{s-1}} \theta_s \int_{\theta_s} f(\theta_s|\theta_{s-1} dH(\theta^{s-1})} \tag{3.6.5}
$$

Use (3.6.1) to obtain

$$
\begin{aligned}
\eta_s(\theta_s) = & \frac{\lambda}{(1+r)^{s-1}} \int_{\theta^{s-1}} \int_{\theta_s}^{\overline{\theta}} dF_s(\tilde{\theta}_s|\theta^{s-1}) dH(\theta^{s-1}) \\
& - \beta^{s-1} \int_{\theta^{s-1}} \int_{\theta_s}^{\overline{\theta}} U'(R_s(\theta^{s-1}, \theta_s)) dF_s(\tilde{\theta}_s|\theta^{s-1}) d\tilde{H}(\theta^{s-1}) \\
& - \int_{\theta^{s-1}} \int_{\theta_s}^{\overline{\theta}} \mu_s(\theta^{s-1}, \theta_s) U''(R_s(\theta^{s-1}, \theta_s)) d\tilde{\theta}_s d\theta^{s-1} \\
& + \beta(1+r)(1-\tau_s) \int_{\theta^{s-1}} \mu_{s-1}(\theta^{s-1}) \int_{\theta_s}^{\overline{\theta}} U''(R_2(\theta^{s-1}, \theta_s)) dF_s(\tilde{\theta}_s|\theta^{s-1}) d\theta^{s-1} \qquad (3.6.6)
\end{aligned}
$$

Inserting (3.6.6) in (3.6.5) yields Proposition 3.3.3

3.6.3 Multiplier Functions μ_t

Use (3.6.3) to obtain, with $SOC_s(\theta^s)$ basically being the second-order condition for savings from the individuals problem:

$$
\begin{aligned}
\mu_s(\theta^s) = & \frac{\frac{\lambda}{(1+r)^{s-1}} \tau_{s+1} h(\theta^{s-1}) + (1-\tau_s)\beta(1+r)\mu_{s-1}(\theta^{s-1}) U''\left(R_s(\theta^s)\right) f_s(\theta_s|\theta^{s-1})}{SOC_s(\theta^s)} \\
& + \frac{(1-\tau_{s+1})(1+r) \int_{\theta_{s+1}} \mu_{s+1}(\theta^s, \theta_{s+1}) U''(R_{s+1}(\theta^s, \theta_{s+1}) d\theta_{s+1}}{SOC_s(\theta^s)}, \qquad (3.6.7)
\end{aligned}
$$

Therefore, let's define some terms that should make notation less burdensome:

$$A_s(\theta^s) = \frac{\frac{\lambda}{(1+r)^{s-1}} \tau_{s+1} h(\theta^{s-1})}{SOC_s}$$

$$B_s(\theta^s) = \frac{(1-\tau_s)\beta(1+r)U''\left(R_s(\theta^s)\right) f_s(\theta_s|\theta^{s-1})}{SOC_s}$$

$$C_s(\theta^s, \theta_{s+1}) = \frac{(1-\tau_{s+1})(1+r)U''(R_{s+1}(\theta^s, \theta_{s+1}))}{SOC_s}$$

then, we can rewrite (3.6.7) as

$$\mu_s(\theta^s) = A_s(\theta^s) + B_s(\theta^s)\mu_{s-1}(\theta^{s-1}) + \int_{\theta_{s+1}} C_s(\theta^s, \theta_{s+1})\mu_{s+1}(\theta^s, \theta_{s+1}) d\theta_{s+1}$$

Or, more concretely for $s = T-2$:

$$\begin{aligned}\mu_{T-2}(\theta^{T-2}) =& A_{T-2}(\theta^{T-2}) + B_{T-2}(\theta^{T-2})\mu_{T-3}(\theta^{T-3}) \\ &+ \int_{\theta_{T-1}} C_{T-2}(\theta^{T-2}, \theta_{T-1})\mu_{T-1}(\theta^{T-2}, \theta_{T-1})d\theta_{T-1}\end{aligned} \tag{3.6.8}$$

For $s = T-1$, we get:

$$\mu_{T-1}(\theta^{T-1}) = A_{T-1}(\theta^{T-1}) + B_{T-1}(\theta^{T-1})\mu_{T-2}(\theta^{T-2}) \tag{3.6.9}$$

Now insert (3.6.9) into (3.6.8). Omitting arguments, this yields:

$$\mu_{T-2} = \frac{A_{T-2} + B_{T-2}\mu_{T-3} + \int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_{\theta_{T-1}} C_{T-2}(\theta^{T-2}, \theta_{T-1})B_{T-1}(\theta^{T-1})d\theta_{T-1}}$$

Now insert this into μ_{T-3}...

$$\mu_{T-3} = A_{T-3} + B_{T-3}\mu_{T-4} + \int_{\theta_{T-2}} C_{T-3}\frac{A_{T-2} + B_{T-2}\mu_{T-3} + \int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2} \tag{3.6.10}$$

yielding

$$\mu_{T-3} = \frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_{\theta_{T-2}} C_{T-3}\frac{A_{T-2}+\int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1 - \int_{\theta_{T-2}} \frac{C_{T-3}B_{T-2}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}} \tag{3.6.11}$$

$$\begin{aligned}\mu_{T-4} =& A_{T-4} + B_{T-4}\mu_{T-5} \\ &\int_{\theta_{T-3}} C_{T-4}\frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_{\theta_{T-2}} C_{T-3}\frac{A_{T-2}+\int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1 - \int_{\theta_{T-2}} \frac{C_{T-3}B_{T-2}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}}d\theta_{T-3}\end{aligned} \tag{3.6.12}$$

Rewrite to obtain

$$\mu_{T-4} = \left[1 - \int_{\theta_{T-3}} C_{T-4}B_{T-3}\left[1 - \int_{\theta_{T-2}} C_{T-3}B_{T-2}\left[1 - \int_{\theta_{T-1}} C_2B_{T-1}d\theta_{T-1}\right]^{-1}\right]^{-1}\right]^{-1}$$

$$A_{T-4} + B_{T-4}\mu_{T-5}$$

$$\int_{\theta_{T-3}} C_{T-4}\frac{A_{T-3} + \int_{\theta_{T-2}} C_{T-3}\frac{A_{T-2}+\int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1 - \int_{\theta_{T-2}} \frac{C_{T-3}B_{T-2}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}}d\theta_{T-3} \tag{3.6.13}$$

Now let's also calculate μ_{T-5} to make sure how the pattern looks like. Now make another definition

$$\mu_{T-5} = \left[1 - \int_{\theta_{T-4}} C_{T-5}B_{T-4}\left[...\left[1 - \int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}\right]^{-1}....\right]^{-1}\right]^{-1}$$

$$A_{T-5} + B_{T-5}\mu_{T-6}$$

$$\int_{\theta_{T-4}} C_{T-5}\frac{A_{T-4} + \int_{\theta_{T-3}} C_{T-4}\frac{A_{T-3}+\int_{\theta_{T-2}} C_{T-3}\frac{A_{T-2}+\int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1-\int_{\theta_{T-2}} \frac{C_{T-3}B_{T-2}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}}}{1 - \int_{\theta_{T-3}} \frac{C_{T-4}B_{T-3}}{1-\int_{\theta_{T-2}} \frac{C_{T-3}B_{T-2}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}}} \tag{3.6.14}$$

Now define

$$D_s = \left[1 - \int_{\theta_{s+1}} C_sB_{s+1}\left[1 - \int_{\theta_{s+2}} C_{s+1}B_{s+2}\left[...\left[1 - \int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}\right]^{-1}....\right]^{-1}d\theta_{s+2}\right]^{-1}d\theta_{s+1}\right]$$

Using this definition, we can write μ_{T-5} as

$$\mu_{T-5} = \frac{A_{T-5} + B_{T-5}\mu_{T-6} + \int_{\theta_{T-4}} C_{T-5}\frac{A_{T-4}+\int_{\theta_{T-3}} C_{T-4}\frac{A_{T-3}+\int_{\theta_{T-2}} C_{T-3}\frac{A_{T-2}+\int_{\theta_{T-1}} C_{T-2}A_{T-1}}{D_{T-2}}}{D_{T-3}}}{D_{T-4}}}{D_{T-5}} \tag{3.6.15}$$

It now turns out helpful to make another definition:

$$E_s = \int_{\theta_{s+1}} C_s \frac{A_{s+1} \int_{\theta_{s+2}} C_{s+1} \frac{A_{s+2} + \int_{\theta_{s+3}} C_{s+2} \frac{A_{s+3} + \int_{\theta_{s+4}} C_{s+3} \frac{A_{s+4} \cdots\cdots}{D_{s+4}}}{D_{s+3}}}{D_{s+2}}}{D_{s+1}}$$

Then we can write μ_{T-5} as

$$\mu_{T-5} = \frac{A_{T-5} + B_{T-5}\mu_{T-6} + E_{T-5}}{D_{T-5}}$$

In genereal, we thus obtain:

$$\mu_s = \frac{A_s + B_s\mu_{s-1} + E_s}{D_s}$$

For the second period, we thus obtain

$$\mu_2 = \frac{A_2 + B_2\mu_1 + E_2}{D_2} \tag{3.6.16}$$

then we should get

$$\mu_1 = \frac{A_1 + E_1}{D_1} \tag{3.6.17}$$

Now we can recursively calculate all other μ_t for $t = 2, ..., T$.

In equation (3.6.17) one can see that the $\mu_1(\theta_1) = 0$ if savings taxes are zero. Recursive calculation reveals that all μ_t are equal to zero.

Using these results one can show that in the T=2 case the Euler parts in the optimal tax formulas are equal to the behavioral responses induced by labor taxes on savings.

3.6.4 The Lagrangian – Age-Independent Taxes

Here we have $y_t(\theta_t) = y(\theta_t)$ and $M_t(\theta_t) = M(\theta_t)$

$$
\begin{aligned}
\mathcal{L} = & \int_{\theta_1} \left(\sum_{t=1}^{T} \beta^{t-1} \int_{\theta^t} U \left(M(\theta_t) - a_{t+1}(\theta^t) + (1-\tau)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y(\theta_t)}{\theta_t}\right) \right) dH(\theta^t) \right) d\tilde{F}_1(\theta_1) \\
& + \lambda \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\theta^{t-1}} \int_{\theta_t} y(\theta_t) - M(\theta_t) + \tau(1+r)a_t(\theta^{t-1}) dF_t(\theta_t|\theta_{t-1}) dH(\theta^{t-1}) \\
& + \sum_{t=1}^{T-1} \int_{\theta^t} \mu_t(\theta^t) \Bigg[U' \left(M(\theta_t) - a_{t+1}(\theta^t) + (1-\tau)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y(\theta_t)}{\theta_t}\right) \right) \\
& - \beta(1+r)(1-\tau_{t+1}) \int_{\theta_{t+1}} U' \Bigg(M(\theta_{t+1}) - a_{t+2}(\theta^t, \theta_{t+1}) \\
& + (1-\tau)(1+r)a_{t+1}(\theta^t) - \Psi\left(\frac{y(\theta_{t+1})}{\theta_{t+1}}\right) \Bigg) dF_{t+1}(\theta_{t+1}|\theta^t) \Bigg] d\theta^t \\
& + \int_{\theta} \eta(\theta) \frac{\partial \left(M(\theta) - \Psi\left(\frac{y(\theta)}{\theta}\right) \right)}{\partial \theta} d\theta - \int_{\theta} \eta(\theta) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{y(\theta)}{\theta^2} d\theta
\end{aligned}
$$

Partially integrating $\int_{\theta} \eta(\theta) \frac{\partial\left(M(\theta)-\Psi\left(\frac{y(\theta)}{\theta}\right)\right)}{\partial\theta} d\theta$ yields

$$
\eta(\overline{\theta}) \left(M(\overline{\theta}) - \Psi\left(\frac{y(\overline{\theta})}{\overline{\theta}}\right) \right) - \eta(\underline{\theta}) \left(M(\underline{\theta}) - \Psi\left(\frac{y(\underline{\theta})}{\underline{\theta}}\right) \right) - \int_{\theta} \eta'(\theta) \left(M(\theta) - \Psi\left(\frac{y(\theta)}{\theta}\right) \right) d\theta
$$

The derivatives with respect to the endpoint conditions yield $\forall t: \ \eta_t(\overline{\theta}) = \eta_t(\underline{\theta}) = 0$.

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial M(\theta)} = & - \sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\theta^{t-1}} f(\theta|\theta_{t-1}) dH(\theta^{t-1}) \\
& + \sum_{t=1}^{T} \beta^{t-1} \int_{\theta^{t-1}} U'(R_t(\theta^{t-1}, \theta)) f(\theta|\theta_{t-1}) d\tilde{H}(\theta^{t-1}) \\
& + \sum_{t=1}^{T-1} \int_{\theta^{t-1}} \mu_t(\theta^{t-1}, \theta) U''(R_t(\theta^{t-1}, \theta)) d\theta^{t-1} \\
& - \sum_{t=2}^{T} \beta(1+r)(1-\tau) \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1}) U''(R_t(\theta^{t-1}, \theta)) f(\theta|\theta^{t-1}) d\theta^{t-1} \\
& - \eta'(\theta) = 0 \qquad (3.6.18)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial y(\theta)} =& \sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\theta^{t-1}} f(\theta|\theta_{t-1}) dH(\theta^{t-1}) \\
&- \sum_{t=1}^{T} \beta^{t-1} \int_{\theta^{t-1}} U'(R_t(\theta^{t-1},\theta)) \Psi' \left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} f(\theta|\theta_{t-1}) d\tilde{H}(\theta^{t-1}) \\
&+ \sum_{t=1}^{T-1} \int_{\theta^{t-1}} \mu_t(\theta^{t-1},\theta) U''(R_t(\theta^{t-1},\theta)) \Psi' \left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} d\theta^{t-1} \\
&- \sum_{t=2}^{T} \beta(1+r)(1-\tau) \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1}) U''(R_t(\theta^{t-1},\theta)) \Psi' \left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} f(\theta|\theta^{t-1}) d\theta^{t-1} \\
&- \eta'(\theta) \Psi' \left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} - \eta(\theta) \left(\Psi' \left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta^2} + \Psi'' \left(\frac{y(\theta)}{\theta}\right) \frac{y(\theta)}{\theta^2} \right) = 0 \qquad (3.6.19)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_{s+1}(\theta^s)} =& \frac{\lambda}{(1+r)^{s-1}} \tau h(\theta^{s-1}) - \mu_s(\theta^s) U''\left(R_s(\theta^s)\right) \\
&- (1-\tau)^2 \beta (1+r)^2 \mu_s(\theta^s) \int_{\theta_{s+1}} U''(R_s(\theta^s,\theta_{s+1})) dF(\theta_{s+1}|\theta^s) \\
&+ (1-\tau)\beta(1+r)\mu_{s-1}(\theta^{s-1}) U''\left(R_s(\theta^s)\right) f(\theta_s|\theta^{s-1}) \\
&+ (1-\tau)(1+r) \int_{\theta_{s+1}} \mu_{s+1}(\theta^s,\theta_{s+1}) U''(R_{s+1}(\theta^s,\theta_{s+1}) d\theta_{s+1} = 0 \qquad (3.6.20)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \tau} =& \sum_{t=2}^{T} \frac{\lambda}{(1+r)^{t-2}} \int_{\theta^{t-1}} a_t(\theta^{t-1}) dH(\theta^{t-1}) \\
&- \sum_{t=2}^{T} \beta^{t-1}(1+r) \int_{\theta^{t-1}} a_t(\theta^{t-1}) \int_{\theta_t} U'(R_t(\theta^{t-1},\theta_t)) dF(\theta_t|\theta_{t-1}) d\tilde{H}(\theta^{t-1}) \\
&- \sum_{t=1}^{T-1} (1+r) \int_{\theta^{t-1}} a_t(\theta^{t-1}) \int_{\theta_t} \mu_t(\theta^{t-1},\theta_t) U''(R_t(\theta^{t-1},\theta_t)) d\theta_t d\theta^{t-1} \\
&+ \sum_{t=2}^{T} \beta(1+r)^2(1-\tau_t) \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta_t} U''(R_t(\theta^{t-1},\theta_t)) a_t(\theta^{t-1}) dF(\theta_t|\theta_{t-1}) d\theta^{t-1} \\
&+ \sum_{t=2}^{T} \beta(1+r) \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta_t} U'(R_t(\theta^{t-1},\theta_t)) dF(\theta_t|\theta_{t-1}) d\theta^{t-1} = 0 \qquad (3.6.21)
\end{aligned}
$$

Dividing (3.6.19) by $\Psi' \frac{1}{\theta}$ and adding (3.6.18) yields

$$\frac{\mathcal{T}'(y(\theta))}{1-\mathcal{T}'(y(\theta))} = \frac{\varepsilon(\theta)+1}{\varepsilon(\theta)} \frac{\eta(\theta)}{\lambda\theta \sum_{t=1}^{T} \int_{\theta_{t-1}} \int_{\theta_t} f(\theta_t|\theta_{t-1} dH(\theta^{t-1})} \tag{3.6.22}$$

Use (3.6.18) to obtain

$$\begin{aligned}
\eta(\theta) = & \sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\theta^{t-1}} \int_{\theta}^{\overline{\theta}} dF(\tilde{\theta}|\theta^{t-1}) dH(\theta^{t-1}) \\
& - \sum_{t=1}^{T} \beta^{t-1} \int_{\theta^{t-1}} \int_{\theta}^{\overline{\theta}} U'(R_t(\theta^{t-1}, \theta_t)) dF(\tilde{\theta}|\theta^{t-1}) d\tilde{H}(\theta^{t-1}) \\
& - \sum_{t=1}^{T} \int_{\theta^{t-1}} \int_{\theta}^{\overline{\theta}} \mu_t(\theta^{t-1}, \tilde{\theta}) U''(R_t(\theta^{t-1}, \tilde{\theta})) d\tilde{\theta} d\theta^{t-1} \\
& + \sum_{t=1}^{T} \beta(1+r)(1-\tau) \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta}^{\overline{\theta}} U''(R_t(\theta^{t-1}, \tilde{\theta})) dF(\tilde{\theta}|\theta^{t-1}) d\theta^{t-1}
\end{aligned} \tag{3.6.23}$$

Inserting (3.6.23) into (3.6.22) yields the formula for optimal labor tax rates.

3.6.5 Details on Numerical Simulations

We use the empirical model from Karhan and Ozkan, who estimate their model using the PSID.

$$y_a^i = f(X_a^i) + \tilde{y}_a^i$$

$$\tilde{y}_a^i = \alpha^i + z_a^i + \phi\epsilon_a^i$$

$$z_a^i = \rho_a z_{a-1}^i + \pi\eta_a^i$$

In this section, let y_a^i be the log income of individual i with age a. Let $f(X_a^i)$ be function of individual observable characteristics – in our case these are education dummies and polynomials in age. The residual $\tilde{y}_a^i$ of the first-stage regression is modeled as a function of an individual fixed-effect α^i, a permanent component z_a^i and a transitory component ϵ_a^i. Crucial are the estimated persistence ρ_a parameter as well as the second moments σ_α^2, σ_α^2, σ_η^2 and σ_ϵ^2, where the latter three are also conditional on the three age groups. We take estimates for all parameters from Karhan and Ozkan. The key parameters are $\rho_Y = 0.88$, $\rho_M = 0.97$ and $\rho_O = 0.96$ for persistence, and $\sigma_{\eta,Y}^2 = 0.027$, $\sigma_{\eta,M}^2 = 0.013$, $\sigma_{\eta,M}^2 = 0.026$, $\sigma_{\epsilon,Y}^2 = 0.059$, $\sigma_{\epsilon,M}^2 = 0.059$, $\sigma_{\epsilon,O}^2 = 0.068$. Finally ϕ and π are loading factors which are allowed to depend on time in Karahan and Ozkan. We set π=1.1253 and $\phi = 1.1115$ corresponding to the values from 1996 and in the roughly in the middle of all estimates for the years from 1968-1997. Simulating millions earnings histories, we treat those as income data and calibrate the respective skill distribution using a flat tax approximation for the US labor tax of 25%.

Chapter 4

Education Policies and Taxation without Commitment [1]

This chapter is joint work with Dominik Sachs.

It is submitted for publication to the *Journal of Public Economics*.

4.1 Introduction

Public finance economists have long recognized that the challenges involved in the design of optimal education policies and income tax systems are intimately related. Income taxation influences the incentives to invest into education.[2] Education subsidies and policies, in turn, influence the choice of an optimal income tax system as they have a direct effect on both the level and the distribution of wages. Many papers have studied the design of education and tax policies jointly from a normative perspective – see, for example,Bovenberg and Jacobs (2005) for a state-of-the-art treatment in a heterogenous agent model.[3] This strand of literature assumes that individuals rationally

[1]Contact: sebastian.findeisen@uzh.ch, dominik.sachs@uni-konstanz.de. We thank Fabrizio Zilibotti and seminar participants in Konstanz and Zurich for helpful comments and discussions. Sebastian Findeisen acknowledges the support of the University of Zurich (Forschungskredit of the University of Zurich, grant no. 53210603).

[2]See Abramitzky and Lavy (2012) for recent quasi-experimental evidence on the negative effect of redistributive taxation on education investment. More structural and model based approaches as the classic work by Trostel (1993) also have found big effects of income taxation on human capital investment.

[3]See Richter (2009) for a recent treatment in a Ramsey setting with a representative agent.

make human capital investment decisions, reacting to incentives set by the tax code and education subsidies. Importantly, income taxes do not change after individuals have made education decisions.

Boadway, Marceau, and Marchand (1996) have drawn attention to the issue of time-consistency, in the spirit of Kydland and Prescott (1977) , inherent in the design of optimal tax and education policies. If the government lacks a device to credibly commit to tax policies at the time individuals make education decisions, this can dramatically depress the incentives of young individuals to invest into human capital. In their framework, they show that this underinvestment arises and make a case for mandatory education as a second-best policy in the presence of commitment problems.

In this paper, we take a fresh look on the implications of limited commitment and policy credibility on education and tax policies. Consistent with real world practices, the government can decide to subsidize different levels of education at different rates. The idea here is that governments typically intervene at primary, secondary and tertiary education levels. However, as we will also exploit in our empirical section, the rate at which these different education levels are subsidized is very different. We formalize this by allowing the government to set a nonlinear schedule of education subsidies. The income tax is linear and the revenue is redistributed lump-sum and used to finance education subsidies. We derive our results in a transparent and simple heterogenous agent model with two types (Stiglitz (1982)) and two periods: an education and a working period. Consistent with empirical evidence, individual wages are determined by both innate abilities and education levels.

We first consider the two polar cases where the government has full commitment to stick to tax promises and no commitment at all. Under full commitment the optimal income tax rate takes into account education incentives. The tax rate is smaller, when the effect of education on wages is large relative to the effect of innate abilities on wages for the initially high skilled. Intuitively, the more important the role of education for wages the more important taxes become to incentivize high-skilled agents to self-select into high education level.

Without any commitment, no tax promise of the government is credible and individuals rationally anticipate that the government re-optimizes after education is sunk. In line with previous results, this leads to excessive taxation and depresses human capital investment. The important result concerns the design of education subsidies. We show they tend to become more progressive when there are commitment problems. The intuition is that a higher subsidy for low types and a lower subsidy for high types will compress the distribution of education. As education inequality is reduced, also the wage distribution in the next period will be more compressed. As wage inequality has decreased, the redistributive government will set a lower tax rate in the second period. This lower tax rate will help to boost education incentives and help to alleviate the commitment problem. This is consistent with the recent results from Farhi, Sleet, Werning, and Yeltekin (2012) who first detected a similar channel for nonlinear capital taxation.

We move on to study intermediate scenarios, where the government has some form of limited commitment, nesting the two polar cases. Under partial commitment, labor income taxes are still designed to take into account their effect on education incentives. However, the strength of the effect is decreasing the more severe the commitment problem is. Too low tax promises lack credibility. Education policies become more progressive the more severe the commitment problem, if the government is sufficiently redistributive towards low types. Another way to look at this result is to interpret the design of education and tax policies as a choice to engage in redistribution *ex-ante* through more progressive education subsidies as opposed to engage in redistribution *ex-post* through income taxation. With a lack of commitment, the government tries to weaken its own temptation to engage in costly ex-post redistribution by increasing the amount of ex-ante redistribution.

We conclude by providing suggestive evidence for the described mechanisms in the form of cross-country correlations. We proxy for commitment power using data from the World Bank‘s Worldwide Governance Indicators. Specifically, we use the variable Government Effectiveness capturing "...the quality of policy formulation and implementation, and the credibility of the government‘s commitment to such policies." (Kaufmann, Kraay, and Mastruzzi (2010)). Controlling for income, geographical variables and the overall share of government involvement in education, we find a robust correlation, indicating that countries with higher policy credibility employ more regressive subsidies.

Despite the mentioned articles by Farhi, Sleet, Werning, and Yeltekin (2012) and Boadway, Marceau, and Marchand (1996), this paper is related to the work on time inconsistency and education policies by Konrad (2001) and Andersson and Konrad (2003).[4] Konrad (2001) shows how the time inconsistency problem is alleviated by the presence of private information in an optimal taxation framework. In particular, he shows that the strong no-education result obtained in Boadway no longer applies, as with private information some rents of education are still captured by individuals, preserving some incentives to invest into education. In our framework, a similar logic applies as the government uses a linear tax income rate together with lump-sums, in the spirit of a simple negative income tax systems. This also preserves some incentives to invest into education, even in the complete absence of credible policy promises, as

[4] In a related paper, Pereira (2009) studies linear education subsidies and shows that it offsets some of the excessive redistribution with income taxes done, when the government lacks commitment. In our framework with nonlinear education policies, as used in the real world, this does typically not arise and we show that if the social welfare function is sufficiently redistributive that a lack of commitment leads to more progressive subsidies.

full equalization of incomes is not feasible. Andersson and Konrad (2003) investigate education policies chosen by extortionary governments lacking commitment and how migration and tax competition affects policies. We depart from these papers by placing our focus on nonlinear education policies, as used in the real world, and how they are affected by commitment issues.

4.2 The Model: Full Commitment and Complete Lack of Commitment

In this section, we characterize government policies under two scenarios. The first scenario is one in which all announcements concerning tax policies of the government are credible. In the second scenario there is a complete lack of commitment and credibility. Individuals will anticipate a government‘s time inconsistency and that it always re-optimize over tax policies once education decisions are sunk.

4.2.1 Environment

We consider a two-period model, where ex-ante heterogeneous agents make an educational investment in period 1. In Period 2, they make a labor leisure decision. More formally, there are two types of ex-ante heterogeneous agents. The θ_1-type and the θ_2-type with $\theta_2 > \theta_1$. Their masses are $f(\theta_1)$ and $f(\theta_2)$ with $f(\theta_1) + f(\theta_2) = 1$. In Period 1, they make a monetary educational investment e. The wage w they earn in period 2 is a function of innate type and education, i.e. $w(\theta, e)$.

We impose three intuitive assumptions on the wage function $w(\theta, e)$. First, education is productive and raises wages $\frac{\partial w(\theta,e)}{\partial e} > 0$. Second education and innate ability are complements implying higher marginal returns to education for the higher innate type: $\frac{\partial w(\theta_2,e)}{\partial e} - \frac{\partial w(\theta_1,e)}{\partial e} > 0$. Finally, innate abilities positively influence wages so $w(\theta_2, e) - w(\theta_1, e) > 0$. None of these assumption are needed for most of the results we derive in the sense that all formulas would be valid if we deviate from those assumptions. The assumption ease the understanding of the model, however, and have strong empirical support.[5]

We assume quasi-linear preferences. To minimize the notational burden we often write all the variables not as a function of θ but with subscript instead. E.g. e_1 instead of $e_1(\theta_1)$ or w_2 instead of $w(e_2(\theta_2), \theta_2)$. The utility functions are $U^1 = c^1$ in period one and

[5]Card (1999) summarizes a long strand of literature estimating the *causal* effect of education on earnings. Carneiro and Heckman (2005) and Lemieux (2006a), among others, document complementarity between innate skills and formal education. Taber (2001) and Hendricks and Schoellman (2012b) suggest that much of the rise in the college premium may be attributed to a rise in the demand for unobserved skills, which are predetermined and independent of education.

$U^2 = c^2 - \Psi(h)$ in period two, where h are hours worked. For simplicity but without loss of generality we assume that Ψ exhibits a constant elasticity of labor supply ϵ, where $\epsilon < 1$.[6] Before tax income is denoted by $y_i = w_i h_i$.

We are considering redistributive linear taxation. That is, we are interested in the policies of a government that is interested in redistributing from the high type θ_2 to the low type θ_1 via linear taxes used to finance a lump-sum rebate such as in a negative income tax system. To capture this redistributive concern, we set the Pareto weights $\tilde{f}(\theta_1)$ and $\tilde{f}(\theta_2)$ such that $\frac{\tilde{f}(\theta_1)}{f(\theta_1)} > \frac{\tilde{f}(\theta_2)}{f(\theta_2)}$. When deciding about the optimal degree of redistribution the government has to take into account that taxes will (i) lower incentives to invest into education and (ii) lower incentives to work in the second period. The education margin, however, can also be influenced by nonlinear education subsidies. Before looking at optimal policies in the limited commitment framework, we look at the simple benchmark case of exogenous education where commitment issues do not arise and then at optimal full commitment policies in the case of endogenous education. We delegate all proofs and derivations to the appendix.

4.2.2 Optimal Policies with Exogenous Education

Assume that education levels e_1 and e_2 are exogenously set. In that case, the only relevant margin for the government when choosing taxes is the labor-leisure margin. The problem of the government then simply is

$$\begin{aligned}\max_t \quad & \tilde{f}(\theta_1)\left((1-t)w(e_1,\theta_1)h(t,w_1) - \Psi[h(t,w_1)]\right) \\ & + (1-\tilde{f}(\theta_1))\left((1-t)w_2(e_2,\theta_2)h(t,w_2) - \Psi[h(t,w_2)]\right) \\ & + t\left(f(\theta_1)w(e_1,\theta_1)h(t,w_1) + (1-f(\theta_1))w(e_2,\theta_2)h(t,w_2)\right), \end{aligned} \tag{4.2.1}$$

where

$$h(t,w_i) = \arg\max_h (1-t)hw_i - \Psi(h).$$

The first two lines of the government problem (4.2.1) is welfare without without the transfer and the third line is the size of the transfer both agents receive. This implies the government budget is already plugged into the welfare function in place of the transfer.

The government thus only has to choose t optimally and thereby take into account that individuals respond their hours worked h. It is then easy to show that the optimal linear tax rate t^{ex} in the case with exogenous human capital satisfies

[6] Chetty (2012b) bounds the labor supply elasticity for reasonable values of adjustment costs to lie between 0.23 and 0.61.

$$\frac{t^{ex}}{1-t^{ex}} = \frac{\left(\tilde{f}(\theta_1) - f(\theta_1)\right)\left(\frac{y_2-y_1}{\bar{y}}\right)}{\varepsilon}, \tag{4.2.2}$$

where $\bar{y}$ is average income $f(\theta_1)y_1 + f(\theta_2)y_2$. The optimal tax rate is increasing in redistributive preferences $\left(\tilde{f}(\theta_1) - f(\theta_1)\right)$, increasing in inequality measured by $\frac{y_2-y_1}{\bar{y}}$ and decreasing in the elasticity of labor supply. The formula (4.2.2) is a variation for the optimal linear tax rate of Sheshinski (1972).[7]

In the next subsection, we will endogenize education decisions and consider education policies as an additional instrument of the government.

4.2.3 Optimal Policies with Endogenous Education and Full Commitment

As already argued, we will now consider the case where the educational decision is endogenous and the government can influence the decision of the agents by setting a nonlinear subsidy schedule. Allowing for nonlinearities adds an incentive constraint to the problem, as in the seminal model with nonlinear labor taxes (Stiglitz 1982). The problem of the government then is

$$\begin{aligned}
\max_{c_1^1, c_2^1 2, t, e_1, e_2} \quad & \tilde{f}(\theta_1)\left((1-t)w(e_1,\theta_1)h(t,w_1) - \Psi[h(t,w_1)]\right) \\
& + (1-\tilde{f}(\theta_1))\left((1-t)w_2(e_2,\theta_2)h(t,w_2) - \Psi[h(t,w_2)]\right) \\
& + t\left(f(\theta_1)w(e_1,\theta_1)h(t,w_1) + (1-f(\theta_1))w(e_2,\theta_2)h(t,w_2)\right) \\
& + (\tilde{f}(\theta_1) - f(\theta_1))c_1^1 + (\tilde{f}(\theta_2) - f(\theta_2))c_2^1 - f(\theta_2)e_2 - f(\theta_1)e_1,
\end{aligned} \tag{4.2.3}$$

subject to the incentive compatibility constraint

$$c_2^1 + (1-t)w_2h(t,w_2) - \Psi(h(t,w_2)) \geq c_1^1 + (1-t)w(e_1,\theta_2)h(t,w(e_1,\theta_2)) - \Psi(h(t,w(e_1,\theta_2))) \tag{4.2.4}$$

with multiplier η and subject to behavioral responses $h(t,w_i) = \arg\max_h (1-t)hw_i - \Psi(h)$.[8]

Period 1 consumption is given by c_i^1 for both agents. One can think of it as student grants offered by the governments, so individuals can finance their living expenses while taking education.[9] Notice that in the incentive constraint (4.2.4) the deviation

[7] See Stantcheva (2013) for a smilar formula in a discrete type setting.

[8] We focus on downward redistributive taxation, where the incentive constraint of the high type is binding.

[9] Instead of first period consumption one could also work with an education dependent transfer in the second period, since with risk-neutrality the timing of consumption across periods is undefined.

utility on the right-hand-side, the terms $w(e_1, \theta_2)$ and $h(t, w(e_1\theta_2))$ show up. A deviating high-skilled agent receives the education level of the low skilled agent e_1. The wage she receives differs, however, because of the effect of innate abilities on wages. To keep notation simple we will call this $w(e_1, \theta_2) = w^c$, with a c for counterfactual as in equilibrium by incentive compatibility the wage will never be observed. We call the associated hour choice $h(t, w(e_1\theta_2)) = h^c$ and associated income $y^c = h^c w^c$.

The problem differs from the problem in Section 4.2.2 as the government now also chooses e_1 and e_2 and that an ex-ante incentive compatibility constraint has to be satisfied. Intuitively, this captures the effect of education and tax policies on human capital investment decisions. We summarize the results of this problem in the following propositions.

Proposition 4.2.1. *In a full commitment economy, the optimal linear tax rate satisfies*

$$\frac{t^f}{1-t^f} = \frac{(\tilde{f}(\theta_1) - f(\theta_1))\left(\frac{y_2 - y_1}{\bar{y}}\right) - \eta\left(\frac{y_2 - y_2^c}{\bar{y}}\right)}{\epsilon},$$

where $y_2^c = w(\theta_2, e_1)h(w(\theta_2, e_1), t^f)$ and the multiplier is equal to $\eta = \tilde{f}(\theta_1) - f(\theta_1)$.

The tax rate with endogenous education decisions is still increasing in income inequality and decreasing in the labor supply elasticity. As can be seen, there is an additional force given by $\eta\left(\frac{y_2 - y_2^c}{\bar{y}}\right)$ in the numerator as compared to the case where education is taken as exogenous.[10] It decreases the optimal tax rate, and the effect is stronger the bigger the difference $y_2 - y_2^c$. y_2^c is the income level the high type θ_2 would attain when only taking the education level of the low type e_1. The difference, hence, captures the effect of a higher education level for the high type on her earnings. The more important the effect of education on earnings, the smaller the tax rate tends to be. In one extreme case, additional education does not change wages at all for the high-type, so $y_2 = y_2^c$. There is no need for the optimal tax rate to take into account education incentives, and the formula collapses to the case with exogenous human capital. In another extreme case $y_1 = y_2^c$, so with the same education level both agents would receive the same wage. This would essentially eliminate agent heterogeneity and the optimal tax rate would be zero in a model without risk.

The following proposition directly follows.

Proposition 4.2.2. *Let e_1^* and e_2^* be the solution to the problem (4.2.3). Then the respective optimal linear tax rate is smaller than the linear tax rate as defined by (4.2.2) for $e_1 = e_1^*$ and $e_2 = e_2^*$, i.e. $t^f(e_1^*, e_2^*) < t^{ex}(e_1^*, e_2^*)$.*

[10]In Findeisen and Sachs (2012) we show that a similar effect arises in a model with uncertainty and when taxes are allowed to condition on education.

That is, the planner uses t as an instrument to set education incentives correctly, where the government treating education as exogenous will typically set labor taxes too high.

Governments do rely on education subsidies to increase the incentives to invest into education. Let us next define the education subsidy s for each θ:

$$(1 - s(\theta_i)) = (1 - t)\frac{\partial w_i}{\partial e} \;\; \forall\, i = 1, 2$$

It directly follows from the first-order conditions of individuals when they face a subsidy schedule such that they only pay $e - S(e)$ for their education and $s(e)$ is the marginal rate. We now characterize the optimal subsidy.

Proposition 4.2.3. *In a full commitment economy, education subsidies satisfy*

$$s^f(\theta_1) = t^f \frac{\partial w_1}{\partial e_1} h_1(1+\epsilon) - \frac{\eta}{f(\theta_1)}(1-t^f)\left[h_2^c \frac{\partial w_2^c}{\partial e_1} - h_1 \frac{\partial w_1}{\partial e_1}\right]$$

and

$$s^f(\theta_2) = t^f \frac{\partial w_2}{\partial e_2} h_2(1+\epsilon).$$

First, looking at the education subsidy for the low type one can see that there are two parts. The first term reflects a fiscal externality term ins the spirit of Bovenberg and Jacobs (2005). The governments offsets the negative effect from labor taxes on the education margin. The larger the labor supply elasticity is, the larger the subsidy. Intuitively, if individuals'working hours react stronger to wage increases, the bigger the fiscal externality on the government budget. Relatedly a bigger the marginal effect of education on the wage and a higher tax rate increase the subsidy. The second term captures the fact innate abilities and education are complements Jacobs and Bovenberg (2011a). The marginal return to education is increasing in innate ability. As the government is redistributive, there is a force towards lowering education subsidies, as they tend to profit more the initially high types. For the high type θ_2 only the fiscal externality part is present as a no-distortion-at-the-top result applies for the second part.

4.2.4 No Commitment

We now characterize policies when the government has no commitment power and contrast them to the full commitment results. We start backwards, looking at optimal tax policies, once education decisions are sunk.

The Problem in Period Two

The problem of the planner is basically equivalent to that of the planner in Section 4.2.2, as the distribution of wages is taken as exogenous. In particular the same tax formula applies:

$$\frac{t^{nc}}{1-t^{nc}} = \frac{\left(\tilde{f}(\theta_1) - f(\theta_1)\right)\left(\frac{y_2-y_1}{\bar{y}}\right)}{\varepsilon}. \tag{4.2.5}$$

We are going to write the optimal tax rate for the second period planner as as function of both education levels, so $t^{nc}(e_1, e_2)$.

The Problem in Period One

In the first period, the planner anticipates that he will set taxes according to (4.2.5). Therefore, in the first period, the problem reads as:

$$\begin{aligned}
\max_{e_1,e_2} = &\ \tilde{f}(\theta_1)\left((1-t^{nc}(e_1,e_2))w_1(e_1,\theta_1)h_1(t^{nc}(e_1,e_2),w_1) - \Psi[h_1(t^{nc}(e_1,e_2),w_1)]\right) \\
&+ (1-\tilde{f}(\theta_1))\left((1-t^{nc}(e_1,e_2))w_2(e_2,\theta_2)h_2(t^{nc}(e_1,e_2),w_2) - \Psi[h_2(t^{nc}(e_1,e_2),w_2)]\right) \\
&+ t^{nc}(e_1,e_2)\left(f(\theta_1)w_1(e_1,\theta_1)h_1(t^{nc}(e_1,e_2),w_1) + (1-f(\theta_1))w_2(e_2,\theta_2)h_2(t^{nc}(e_1,e_2),w_2)\right) \\
&+ (\tilde{f}(\theta_1) - f(\theta_1))c_1^1 + (\tilde{f}(\theta_2) - f(\theta_2))c_2^1 - f(\theta_2)e_2 - f(\theta_1)e_1
\end{aligned}$$

subject to

$$c_2^1 + (1-t^{nc}(e_1,e_2))w_2h_2 - \Psi(h_2) \geq c_1^1 + (1-t^{nc}(e_1,e_2))w_2^c h_2^c - \Psi(h_2^c)$$

One can show the following proposition characterizing education policies without commitment.

Proposition 4.2.4. *In a no commitment economy, education subsidies satisfy*

$$s^{nc}(\theta_1) = t^{nc}\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - \frac{\eta}{f(\theta_1)}(1-t^{nc})\left[h_2^c\frac{\partial w_2^c}{\partial e_1} - h_1\frac{\partial w_1}{\partial e_1}\right] - \frac{\eta}{f(\theta_1)}\frac{\partial t^{nc}}{\partial e_1}(y_2 - y_2^c)$$

and

$$s^{nc}(\theta_2) = t^{nc}\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - \frac{\eta}{f(\theta_1)}\frac{\partial t^{nc}}{\partial e_2}(y_2 - y_2^c)$$

where $\frac{\partial t^{nc}}{\partial e_1} < 0$ *and* $\frac{\partial t^{nc}}{\partial e_2} > 0$ *and the multiplier is again given by* $\eta = \tilde{f}(\theta_1) - f(\theta_1)$.

In comparison to Proposition 4.2.3, the education subsidy for the low type is upward distorted and downward distorted for the high type. Together this tends to make education policies more progressive. The intuition behind this result is clear and simple. In the case of the low type the adjustment is given by:

$$-\frac{\eta}{f(\theta_1)}\frac{\partial t^{nc}}{\partial e_1}(y_2 - y_2^c) > 0$$

A higher education subsidy and level of education for the low type will decrease the optimal tax rate chosen by the government in period two $\frac{\partial t^{nc}}{\partial e_1} < 0$. This will strengthen education incentives. Put differently, the government anticipates its temptation to set too high taxes in the second period. By compressing the distribution of eduction across the two agents, it can avoid some of the harmful spillover from too high taxes on the education margin. Consistent with that argument, there is a downward adjustment in the optimal subsidy for the high type

$$-\frac{\eta}{f(\theta_1)}\frac{\partial t^{nc}}{\partial e_2}(y_2 - y_2^c) < 0,$$

as a higher education level for the high type will tend to increase taxes because of higher income inequality.

4.3 Varying the Degree of Commitment

In the previous section we studied two polar cases. We now look at economies, where the degree of commitment power of the government is allowed to differ, nesting the two cases from the previous section. This allows us to show that smoother versions of our previous results hold.

4.3.1 Costs of Deviating and the Commitment Technology

Following Farhi, Sleet, Werning, and Yeltekin (2012), we introduce an additional credibility constraint on the government problem. It takes the form:

$$\mathcal{W}^2_{PC}(e_1, e_2, t) \geq \mathcal{W}^2_{Dev}(e_1, e_2) - \kappa, \tag{4.3.1}$$

where $\mathcal{W}^2_{PC}(e_1, e_2, t)$ is second period welfare as a function of education levels for both types and the promised tax rate t, under the assumption that the government sticks to its promise. $\mathcal{W}^2_{Dev}(e_1, e_2)$ on the other side of the inequality is the welfare obtained, if the government reneges on its tax promise and effectively takes the distribution of education as exogenous as in Section 4.2.2. If the government reneges, however, and re-optimizes the tax rate, we assume it will incur a output loss of κ.[11] This reduced form allows to flexibly capture forms of *limited commitment*. At the one extreme end, when κ is zero there is no way for the government to credibly commit and we arrive at the case from Section 4.2.4. At the other extreme end, when κ is above some positive threshold $\bar{\kappa} > 0$, all tax promises are fully credible, and we arrive at the full commitment solution, which naturally achieves the highest welfare. In this section we focus on the intermediate cases where κ lies between zero and $\bar{\kappa}$.

4.3.2 Optimal Policies and Discussion

The full problem of a limited commitment economy is stated in the appendix. The following propositions characterize optimal policies for this case.

Proposition 4.3.1. *In a partial commitment economy, the optimal linear tax rate satisfies:*

$$\frac{t^{pc}}{1-t^{pc}} = \frac{(\tilde{f}(\theta_1) - f(\theta_1))\left(\frac{y_2-y_1}{\bar{y}}\right) - \frac{\eta}{1+\zeta}\left(\frac{y_2-y_2^c}{\bar{y}}\right)}{\epsilon},$$

where ζ is the multiplier on the credibility constraint (4.3.1).

One can see how this case nests the full commitment case and the optimal rate from Proposition 4.2.1. If the credibility constraint is not binding for sufficiently high κ, ζ is equal to zero and the government is able to implement the full commitment tax rate. As discussed above, the second term in the numerator reflects how labor taxes are adjusted to provide education incentives and complement education subsidies. This effect is now scaled down by $\frac{1}{1+\zeta}$. The more severe the commitment problem, the bigger ζ tends to be. This will make any tax promises less credible and, anticipating this, the government will set a higher, more credible tax rate.

Next, we characterize the resulting education subsidies.

Proposition 4.3.2. *In a partial commitment economy, education subsidies satisfy:*

$$s^{pc}(\theta_1) = t^{pc}\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - \frac{\eta}{f(\theta_1)}(1-t^{pc})\left[h_2^c\frac{\partial w_2^c}{\partial e_1} - h_1\frac{\partial w_1}{\partial e_1}\right] + \frac{\zeta}{f(\theta_1)}\left(\frac{\partial \mathcal{W}_{PC}}{\partial e_1} - \frac{\partial \mathcal{W}_{Dev}}{\partial e_1}\right)$$

and

$$s^{pc}(\theta_2) = t^{pc}\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) + \frac{\zeta}{f(\theta_2)}\left(\frac{\partial \mathcal{W}_{PC}}{\partial e_2} - \frac{\partial \mathcal{W}_{Dev}}{\partial e_2}\right),$$

where ζ is the multiplier on the credibility constraint (4.3.1).

[11] Farhi, Sleet, Werning, and Yeltekin (2012) show how to microfound such an output loss in a dynamic repeated game, where a deviation today brings a reputational cost borne in the future, because of depressed investment of future generations.

Whenever the credibility constraint is binding, the subsidies get adjusted by $\frac{\zeta}{f(\theta_1)}\left(\frac{\partial \mathcal{W}^2_{PC}}{\partial e_1}-\frac{\partial \mathcal{W}^2_{Dev}}{\partial e_1}\right)$ and $\frac{\zeta}{f(\theta_2)}\left(\frac{\partial \mathcal{W}^2_{PC}}{\partial e_2}-\frac{\partial \mathcal{W}^2_{Dev}}{\partial e_2}\right)$. Economic intuition suggests that the for the low type the force leads to higher education subsidies, and for the high type to lower education subsidies. Whenever there is a commitment problem, the marginal value of low level education goes up as it strengthens the credibility of tax promises. Relatedly the marginal value of high level education goes down, as it increases the temptation to renege on tax promises and increase the tax rate to redistribute. Taken together, a more compressed education distribution leads to a more compressed wage distribution, decreasing the value of a (ex-ante) harmful deviation of the government. With limited commitment, the government wants to avoid excessive ex-post redistribution, by engaging already in ex-ante redistribution through the use of education policies.

The next proposition states more precisely how this intuition shows up formally in our simple model.

Proposition 4.3.3. *For any partial commitment economy, there exists a Pareto weight $\tilde{f}_t(\theta_1)$, s.t. for all $\tilde{f}(\theta_1) \geq \tilde{f}_t(\theta_1)$, the additional distortion for the low type can be signed:*

$$\left(\frac{\partial \mathcal{W}^2_{PC}}{\partial e_1}-\frac{\partial \mathcal{W}^2_{Dev}}{\partial e_1}\right) \geq 0.$$

The highest possible value for any economy for the threshold weight is $\tilde{f}_t(\theta_1) = (1+\epsilon)f(\theta_1)$. Analgously for the high type, for any partial commitment economy, there exists a Pareto weight $\tilde{f}_t(\theta_2)$, s.t. for all $\tilde{f}(\theta_2) \leq \tilde{f}_t(\theta_2)$, the additional distortion can be signed

$$\left(\frac{\partial \mathcal{W}^2_{PC}}{\partial e_2}-\frac{\partial \mathcal{W}^2_{Dev}}{\partial e_2}\right) \leq 0.$$

The lowest possible value for any economy for the threshold weight is $\tilde{f}_t(\theta_2) = 0$, so the Rawlsian case.

Taken together, the strength of the analytical result hinges on how redistributive the government is. A sufficient condition for the upward distortion of the low type is a welfare weight fulfilling $\frac{\tilde{f}(\theta_1)}{f(\theta_1)} > 1+\epsilon$. For realistic values of the labor supply elasticity this implies the condition is fulfilled for most redistributive weights. A sufficient condition for the downward distortion of the high type is the Rawlsian case. In the appendix we present a tighter bound for the welfare weight $\tilde{f}_t(\theta_2)$, suggesting the result is robust to less redistributive social welfare functions. This will now be confirmed in a numerical example illustrating the analytical results.

4.3.3 Numerical Illustration

We assume a equal mass of high and low types $f(\theta_1) = f(\theta_2) = 0.5$ and set the welfare weights to $\tilde{f}(\theta_1) = 0.9$ and $\tilde{f}(\theta_2) = 0.1$. We set the labor supply elasticity to 0.25. For wages, we assume that they are determined by a simple Cobb Douglas production function $w_i = \theta_i^{0.5} e_i^{0.5}$ with equal weights. There is a constant marginal cost of education. We start by assuming that the government has limited commitment power and the cost of reneging on tax promises κ is set at 5% of average output $\bar{y}$, calculated from the full commitment economy.

The equilibrium tax rate is $t^{pc} = 35.56\%$. For comparison it is $t^{fc} = 19.56\%$ in the full commitment case. This illustrates the workings of the formula in Proposition 4.3.1, as the human capital effect on taxes is scaled down relative to the full commitment benchmark, because the government lacks full credibility. A deviating government which would take the wage distribution as exogenously given (Section 4.2.2) would set a tax rate of 65.95%. So although the government lacks commitment power, it still takes human capital investment incentives into account, as it sets a significantly smaller tax rate by about 30 percentage points.

The main predictions from our analysis of education subsides concern the degree of the progressivity of education subsidies, see Propositions 4.3.2 and 4.3.3. In line with these predictions we find that the ratio of the subsidies $\frac{s^{pc}(\theta_2)}{s^{pc}(\theta_1)}$ of the high relative to the low type is 0.99, as compared to 2.38 in the full commitment case. A higher $\frac{s^{pc}(\theta_2)}{s^{pc}(\theta_1)}$ ratio implies a more regressive incidence of subsidies. It becomes clear how the planner choose a more progressive subsidy structure without full commitment, thereby redistributing more resources ex-ante in the first periods, weakening her temptation to engage in costly ex-post redistribution in the second period with excessive taxation.

Finally, we illustrate how the regressivity of education policies varies with the commitment technology. Figure 4.1 plots r $\frac{s^{pc}(\theta_2)}{s^{pc}(\theta_1)}$ against κ as it varies, measured in percentage points of output lost when reneging on tax promises. Moving from left to right, the commitment power of the government gradually increases. In line with the mechanism outlined above, with more commitment the government can afford to set less progressive subsidies, as its credibility increases.

4.4 Empirical Implications

The model predicts a more regressive incidence of subsidies when the ability of a government to commit is high. We now provide suggestive cross-country evidence for this. The estimates of this section should be interpreted as correlations only, of course.

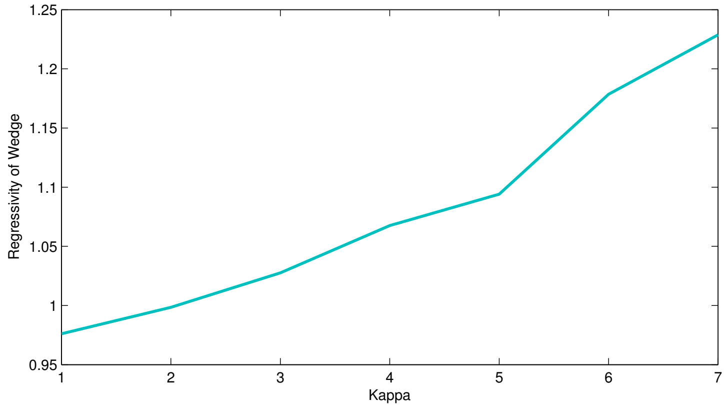


Figure 4.1: Regressivity Education Subsidies

Data. As a measure for commitment power and the credibility of policy announcements we use data from the World Bank's Worldwide Governance Indicators database. We use the variable Government Effectiveness capturing "...the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies." (Kaufmann, Kray and Mastruzzi 2010). To proxy for the regressivity/progressivity of public education subsidies, we use the share of public education expenditures at each educational level relative to total public education expenditures. We then take the share spent on tertiary education relative to the total spending share on all lower levels of education (primary and secondary). The bigger the value of this variable, the more regressive is the incidence of public education expenditure, in the sense that more is spent on tertiary education relative to lower tier education. To construct the measure, we take data from the UNECSO on public educational expenditures across education levels. We also use GDP data, which we take from the World Bank database. We use the year 2008 as the most recent year with a reasonable number of observations. We are left with a sample of 54 countries, for which all the relevant data is available.

Results. For Figure 4.2, we regress our measure of the regressivity of public education on the government credibility index. The correlation is positive and highly significant.[12] The coefficient from column one of Table 1 implies that a one standard deviation increase in policy credibility increases the regressivity of public education expenditures by 0.53 standard deviations. Next, we include continent dummies. Only exploiting the variation within continents increases the credibility coefficient. Adding the log of per capita GDP does not affect the conclusion. Maybe surprisingly, income per capita seems not to

[12]It is even stronger when excluding Lesotho and Cuba, which are two outliers with high regressivity but weak institutional commitment. On the other side, Singapore is an outlier with very strong policy credibility and a high incidence of regressivity.

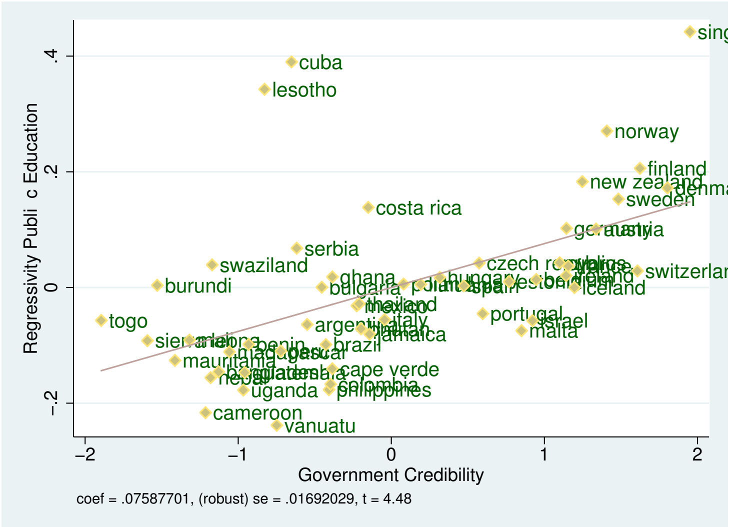


Figure 4.2: Government Education Expenditures and Policy Credibility

be correlated with a more regressive incidence of public expenditure, as is seen in column three. The raw correlation between GDP per capita and our regressivity index is, however, positive and significant (0.49). But as the estimates indicate, this effect vanishes with continent dummies and controlling for government credibility. Finally, we control for the overall share of public education expenditures aggregated across all levels as a fraction of GDP. This approximates for the overall importance of the public sector in providing and paying for education. The main correlation concerning the effect of governmental policy credibility remains unaffected. As column four shows, countries in which the government has a relatively larger stake in education, tend to spend more on higher education.

4.5 Conclusion

Optimal income tax and education policies depend on the degree of commitment power or policy credibility the government has. We build a transparent and simple heterogenous agent model to understand the economic mechanisms involved. Individual wages are determined by both innate abilities and education levels. Without any commitment, the labor income tax does not take into account the incentives to acquire education. When some or full commitment is available, income tax rates are adjusted to incentivize education. The tax rate is smaller, when the effect of education on wages is large relative to the effect of innate abilities on wages for the initially high skilled. We allow the government to subsidize different levels of education at different rates. The main implication of limited commitment is that education policies become more progressive relative to the full commitment benchmark: the government takes into account that a

Table 4.1: Credibility of Government and Education Regressivity

	Dependent variable: Regressivity Education Expenditure			
Policy Credibility	0.759***	1.226***	1.366***	1.252**
	(0.169)	(0.330)	(0.550)	(0.550)
Log GDP			- 0.020	- 0.021
			(0.030)	(0.032)
Total Education				0.031**
Expenditure Share				(0.013)
Continent Dummies	No	Yes	Yes	Yes
R-squared	0.290	0.352	0.356	0.490

Observations: 54. Year: 2008. List of countries see Appendix. Policy Credibility coefficient multiplied by 10. Robust errors. Last column based on 52 observation, since data on Bhutan and Uganda is misssing.

more compressed wage distribution limits its own temptation to tax excessively. By adjusting the distribution of education, the government effectively creates its own commitment device. This mirrors previous findings from Farhi, Sleet, Werning, and Yeltekin (2012) concerning the design of capital taxes. Using data on the credibility of policy announcements from the World Bank database, we find a positive and significant correlation between the degree of commitment power and the how regressive education expenditures are across countries, consistent with the mechanism highlighted in the paper. This correlation is conditional on income and geographical controls.

4.6 Appendix – Chapter Four

4.6.1 Full Commitment Planner

The firs-order conditions of (4.2.3) are

$$\frac{\partial ...}{\partial c_1} = \tilde{f}_1 - f_1 - \eta = 0$$

$$\frac{\partial ...}{\partial c_2} = \tilde{f}_2 - f_2 + \eta = 0$$

$$\begin{aligned}\frac{\partial ...}{\partial t} =& - \tilde{f}(\theta_1)y_1 - (1 - \tilde{f}(\theta_1))y_2 + f(\theta_1)y_1 + (1 - f(\theta_1))y_2 + tf(\theta_1)w_1\frac{\partial h_1}{\partial t} + t(1 - f(\theta_1))w_2\frac{\partial h_2}{\partial t} \\ & - \eta(w_2h_2 - w_2^c h_2^c) = 0\end{aligned}$$

$$\frac{\partial ...}{\partial e_1} = \tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t^F)h_2^c\frac{\partial w_2^c}{\partial e_1}\right] = 0,$$

$$\frac{\partial ...}{\partial e_2} = \tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \eta(1-t)\frac{\partial w_2}{\partial e}h_2 = 0.$$

Optimal tax rate and education policies directly follow from these first-order conditons.

4.6.2 The No Commitment Planner

The FOC for e_1 and e_2

For the first-order condition for e_1 and e_2, we know that their impact on Period 2 welfare via t is zero due to the envelope theorem. Thus the first-order conditions read as:

$$\frac{\partial ...}{\partial e_1} = \tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t^F)h_2^c\frac{\partial w_2^c}{\partial e_1}\right] - \eta\frac{\partial t}{\partial e_1}\left[w_2h_2 - w_c^c h_2^c\right] = 0,$$

$$\frac{\partial ...}{\partial e_2} = \tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \eta(1-t)\frac{\partial w_2}{\partial e}h_2 - \eta\frac{\partial t}{\partial e_2}\left[w_2h_2 - w_c^c h_2^c\right] = 0,$$

The optimal tax rate and education policies directly follow from these first-order conditions.

The derivatives of t with respect to e_1 and e_2

We know that

$$\frac{t}{1-t} = \frac{(\tilde{f}(\theta_1) - f(\theta_1))\left[\frac{w_2h_2 - w_1h_1}{\overline{wh}}\right]}{\epsilon}.$$

We now show that t is increasing in e_2 and decreasing in e_1. Define the implicit function(s): $F(e_i, t(e_i)) = \frac{t}{1-t} - RHS = 0$. Differentiating F w.r.t to e_1 one gets:

$$\begin{aligned}
&\frac{\partial t}{\partial e_1}\left[\frac{1}{(1-t)^2} - \frac{(\tilde{f}(\theta_1) - f(\theta_1))}{\epsilon\overline{y}^2}\left(\left(\frac{\partial y_2}{\partial t} - \frac{\partial y_1}{\partial t}\right)\overline{y} + \left(f(\theta_1)\frac{\partial y_1}{\partial t} + f(\theta_2)\frac{\partial y_2}{\partial t}\right)(y_2 - y_1)\right)\right] \\
&+ \frac{(\tilde{f}(\theta_1) - f(\theta_1))}{\epsilon\overline{y}^2}\frac{\partial y_1}{\partial e_1}\left((y_2 - y_1)f(\theta_1) + \bar{y}\right) \\
&= \frac{\partial t}{\partial e_1}\left[\frac{1}{(1-t)^2} - \frac{(\tilde{f}(\theta_1) - f(\theta_1))}{\epsilon\overline{y}^2}\left(\left(\frac{\partial y_2}{\partial t} - \frac{\partial y_1}{\partial t}\right)\overline{y} + \left(f(\theta_1)\frac{\partial y_1}{\partial t} + f(\theta_2)\frac{\partial y_2}{\partial t}\right)(y_2 - y_1)\right)\right] \\
&+ \frac{(\tilde{f}(\theta_1) - f(\theta_1))}{\epsilon\overline{y}^2}\frac{\partial y_1}{\partial e_1}y_2
\end{aligned}$$

For the second term in the first line we obtain:

$$\begin{aligned}
&\left(\frac{\partial y_2}{\partial t} - \frac{\partial y_1}{\partial t}\right)\overline{y} - \left(f(\theta_1)\frac{\partial y_1}{\partial t} + f(\theta_2)\frac{\partial y_2}{\partial t}\right)(y_2 - y_1) \\
&= \frac{\partial y_1}{\partial t}(-f(\theta_1)y_1 - (1 - f(\theta_1))y_2 - f(\theta_1)y_2 + f(\theta_1)y_1) + \frac{\partial y_2}{\partial t}(f(\theta_1)y_1 \\
&+ (1 - f(\theta_1))y_2 - (1 - f(\theta_1))y_2 + (1 - f(\theta_1))y_1)) \\
&= \frac{\partial y_1}{\partial t}(-y_2) + \frac{\partial y_2}{\partial t}y_1 = \frac{\partial y_1}{\partial 1-t}y_2 - \frac{\partial y_2}{\partial 1-t}y_1 = \varepsilon\frac{y_1y_2}{1-t} - \varepsilon\frac{y_1y_2}{1-t} = 0,
\end{aligned}$$

so it follows $\frac{\partial t}{\partial e_1} < 0$. Similar reasoning shows $\frac{\partial t}{\partial e_2} > 0$

4.6.3 The Partial Commitment Planner

First-Order Conditions

For c_1 and c_2 we get the same FOC as in the full commitment case. For t we get:

$$\begin{aligned}
&\frac{\partial ...}{\partial t} = \\
&(1+\zeta)\left[-\tilde{f}(\theta_1)y_1 - (1 - \tilde{f}(\theta_1))y_2 + f(\theta_1)y_1 + (1 - f(\theta_1))y_2 + tf(\theta_1)w_1\frac{\partial h_1}{\partial t} + t(1 - f(\theta_1))w_2\frac{\partial h_2}{\partial t}\right] \\
&- \eta(w_2h_2 - w_2^ch_2^c) = 0
\end{aligned}$$

$$\frac{\partial ...}{\partial e_1} = \tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t^F)h_2^c\frac{\partial w_2^c}{\partial e_1}\right]$$
$$+ \zeta\frac{\partial w_1}{\partial e_1}\left[\frac{\tilde{f}_1}{f_1}\left((1-t^{PC})h_1(e_1,t^{PC}) - (1-t^{Dev})h_1(e_1,t^{Dev})\right)\right.$$
$$\left. + \left(t^{PC}h_i(e_1,t^{PC}) - t^{Dev}h_1(e_1,t^{Dev})\right)(1+\varepsilon_{h,w})\right] = 0$$

$$\frac{\partial ...}{\partial e_2} = \tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \eta(1-t)\frac{\partial w_2}{\partial e}h_2$$
$$+ \zeta\frac{\partial w_2}{\partial e_2}\left[\frac{\tilde{f}_2}{f_2}\left((1-t^{PC})h_2(e_2,t^{PC}) - (1-t^{Dev})h_2(e_2,t^{Dev})\right)\right.$$
$$\left. + \left(t^{PC}h_2(e_2,t^{PC}) - t^{Dev}h_1(e_2,t^{Dev})\right)(1+\varepsilon_{h,w})\right] = 0$$

And finally, of course,

$$c_2 + w_2h_2 - \Psi(h_2) - c_1 - w_2^c h_2^c + \Psi(h_2^c) = 0$$

and

$$\mathcal{W}_{PC}(e_1,e_2,t) + f(\theta_1)e_1 + f(\theta_2)e_2 + \eta(...) - \mathcal{W}_{Dev}(e_1,e_2,t(e_1,e_2)) = \kappa$$

Proof of Proposition 4.3.3

For the additional term, we immediately get using the envelope condition:

$$\zeta\frac{\partial w_i}{\partial e_i}\left[\frac{\tilde{f}_i}{f_i}\left((1-t^{PC})h_i(e_i,t^{PC}) - (1-t^{Dev})h_i(e_i,t^{Dev})\right)\right.$$
$$\left. + \left(t^{PC}h_i(e_i,t^{PC}) - t^{Dev}h_i(e_i,t^{Dev})\right)(1+\varepsilon_{h,w})\right] \tag{4.6.1}$$

We know that $\frac{\tilde{f}(\theta_1)}{f(\theta_1)} > 1$ and $\frac{\tilde{f}(\theta_2)}{f(\theta_2)} < 1$. In what follows we will write RF_i for $\frac{\tilde{f}_i}{f_i}$ to denote the relative Pareto weight and save on notation. We also simplify the notation for h and write $h_i(e_i,t^{Dev}) = h_i^{dev}$ and similarly for the other expressions.

(4.6.1) can be rearranged as:

$$\left(h_i^{pc} - h_i^{dev}\right) - \left[t^{dev} h_i^{dev} - t^{pc} h_i^{pc}\right] \left[\frac{1 + \varepsilon_{h,w}}{RF_i} - 1\right]. \tag{4.6.2}$$

Notice that $h_i^{pc} - h_i^{dev} > 0$, as the tax rate is always bigger when deviating. Also notice that $t^{dev} h_i^{dev} - t^{pc} h_i^{pc} > 0$, under the assumption that the labor supply elasticity is smaller than one.

We start and show under which conditions there is an upward distortion for the low type. A sufficient condition for equation (4.6.2) to be positive, is:

$$\frac{1 + \varepsilon_{h,w}}{RF_1} - 1 < 0.$$

This proofs the statement made in the proposition.

Next consider the high type. A sufficient condition for equation (4.6.2) to be negative, is a Rawslian welfare weight of $RF_2 = 0$, as can be easily seen. Less redistributive motives may suffice, however. Rewrite (4.6.2):

$$\frac{\left(h_i^{pc} - h_i^{dev}\right)}{\left[t^{dev} h_i^{dev} - t^{pc} h_i^{pc}\right]} \leq \left[\frac{1 + \varepsilon_{h,w}}{RF_i} - 1\right] \tag{4.6.3}$$

for i=2. If RF_2 fulfills this inequality, there is a (weak) downward distortion for the high type. From the set of tax rates t^{dev} and t^{pc}, which solve the respective government problems for given primitives in the economy, there is an upper bound on the left hand side. Denote this upper bound by x, where

$$x = \max_{j \in \Omega} \frac{\left(h_i^{pc}(j) - h_i^{dev}(j)\right)}{\left[t^{dev}(j) h_i^{dev}(j) - t^{pc}(j) h_i^{pc}(j)\right]} > 0$$

and Ω is the set of economies for the different primitives, namely wage functions $w(e, \theta)$, distributions of θ and the labor supply elasticity and welfare weights. The lower bound for RF_2 then comes from:

$$\left[\frac{1 + \varepsilon_{h,w}}{RF_2} - 1\right] > x,$$

or $RF_2 < \frac{1+\varepsilon_{h,w}}{1+x}$, which is always true, for example, if $\varepsilon_{h,w} > x$, since we consider redistributive economies with $RF_2 < 1$.

4.6.4 List of Countries in the Empirical Part

Argentina, Austria, Bangladesh, Belgium, Benin, Bhutan, Brazil, Bulgaria, Burundi, Cameroon, Cape Verde, Colombia, Costa Rica, Cuba, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ghana, Guatemala, , Hungary, Iceland, Ireland, Israel, Italy, Jamaica, Lesotho, Lithuania, Madagascar, Mali, Malta, Mauritania, Mexico, Nepal, New Zealand, Norway, Peru, Philippines, Poland, Portugal, Serbia, Sierra Leone, Singapore, Spain, Swaziland, Sweden, Switzerland, Thailand, Togo, Uganda.

Bibliography

ABRAMITZKY, R., AND V. LAVY (2012): "How Responsive is Investment in Schooling to Changes in Redistributive Policies and in Returns?," .

ACEMOGLU, D., AND D. AUTOR (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics Vol. 4*, ed. by O. Ashenfelter, and D. Card, pp. 1043–1171. Elsevier.

AIYAGARI, S. R. (1994): "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 109(3), 659–684.

AMITI, M., AND D. R. DAVIS (2012): "Trade, Firms, and Wages: Theory and Evidence," *Review of Economic Studies*, 79(1), 1–36.

ANDERBERG, D. (2009): "Optimal Policy and the Risk Properties of Human Capital Reconsidered," *Journal of Public Economics*, 93(9-10), 1017–1026.

ANDERBERG, D., AND F. ANDERSSON (2003): "Investments in Human Capital, Wage Uncertainty, and Public Policy," *Journal of Public Economics*, 87(7-8), 1521 – 1537.

ANDERSSON, F., AND K. A. KONRAD (2003): "Human capital investment and globalization in extortionary states," *Journal of Public Economics*, 87(7), 1539–1555.

ANTRAS, P., L. GARICANO, AND E. ROSSI-HANSBERG (2006): "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics*, 121(1), 31–77.

ARTUC, E., S. CHAUDHURI, AND J. MCLAREN (2010): "Trade Shocks and Labor Adjustments: A Structural Empirical Approach," *American Economic Review*, 100(3), 1008–1045.

ATKINSON, A., AND J. STIGLITZ (1976): "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, 6(1-2), 55–75.

ATKINSON, A. B., T. PIKETTY, AND E. SAEZ (2011): "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 49(1), 3–71.

AUTOR, D. H., D. DORN, AND G. H. HANSON (2012): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," NBER Working Paper No. 18054.

AUTOR, D. H., D. DORN, G. H. HANSON, AND J. SONG (2012): "Trade adjustment: Worker level evidence," Working paper presented at the 2012 SOLE Annual Meetings.

BANKS, J., AND P. DIAMOND (2011): "The Base for Direct Taxation," in *Mirrlees Review: Tax Reform for the 21st century*, chap. 6.

BASTANI, S., S. BLOMQUIST, AND L. MICHELETTO (2011): "The welfare gains of age related optimal income taxation," *International Economic Review*.

BAUM, C. F., M. E. SCHAFFER, AND S. STILLMAN (2007): "Enhanced routines for instrumental variables / generalized method of moments estimation and testing," *Stata Journal*, 7(4), 465–506.

BERNARD, A. B., J. B. JENSEN, AND P. K. SCHOTT (2006): "Survival of the best fit: Exposure to low-wage countries and the (uneven) growth of U.S. manufacturing plants," *Journal of International Economics*, 68(1), 219–237.

BERTOLA, G. (2000): "Labor Markets in the European Union," ifo Studien 1/2000, pp. 99-122.

BEST, M., AND H. KLEVEN (2013): "Optimal Income Taxation with Career Effects of Work Effort," .

BLOOM, N., M. DRACA, AND J. VAN REENEN (2011): "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity," NBER Working Paper No. 16717.

BLOSSFELD, H.-P. (1987): "Labor Market Entry and the Sexual Segregation of Careers in the Federal Republic of Germany," *American Journal of Sociology*, 93(1), 89–118.

BLUM, B. S. (2008): "Trade, technology, and the rise of the service sector: The effects on US wage inequality," *Journal of International Economics*, 74(2), 441–458.

BOADWAY, R., N. MARCEAU, AND M. MARCHAND (1996): "Investment in education and the time inconsistency of redistributive tax policy," *Economica*, pp. 171–189.

BOHACEK, R., AND M. KAPICKA (2008): "Optimal Human Capital Policies," *Journal of Monetary Economics*, 55(1), 1–16.

BOVENBERG, L., AND B. JACOBS (2005): "Redistribution and Education Subsidies are Siamese Twins," *Journal of Public Economics*, 89(11-12), 2005–2035.

BRUELLHART, M., C. CARRERE, AND F. TRIONFETTI (2012): "How wages and employment adjust to trade liberalization: Quasi-experimental evidence from Austria," *Journal of International Economics*, 86(1), 68–81.

CARD, D. (1999): "The causal effect of education on earnings," *Handbook of labor economics*, 3, 1801–1863.

CARNEIRO, P., AND J. HECKMAN (2005): " Human Capital Policies," in *Inequality in America: What Role for Human Capital Policies*, ed. by J. Heckman, and A. Krueger, pp. 77–240. The MIT Press.

CHAPMAN, B. (2006): "Income contingent loans for higher education: International reforms," *Handbook of the Economics of Education*, 2, 1435–1503.

CHEN, S. (2008): "Estimating the Variance of Wages in the Presence of Selection and Unobserved Heterogeneity," *Review of Economics and Statistics*, 90(2), 275–289.

CHETTY, R. (2006): "A new method of estimating risk aversion," *The American Economic Review*, 96(5), 1821–1834.

——— (2012a): "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80(3), 969–1018.

——— (2012b): "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80(3), 969–1018.

CHIQUIAR, D. (2008): "Globalization, Regional Wage Differentials, and the Stolper-Samuelson Theorem: Evidence from Mexico," *Journal of International Economics*, 74(1), 70–93.

CONESA, J. C., S. KITAO, AND D. KRUEGER (2009): "Taxing Capital? Not a Bad Idea after All!," *The American Economic Review*, pp. 25–48.

COURTY, P., AND H. LI (2000): "Sequential Screening," *The Review of Economic Studies*, 67(4), 697–717.

CUNHA, F., AND J. HECKMAN (2007): "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 14(6), 870–893.

——— (2008): "The Evolution of Labor Earnings Risk in the US Economy," *Working Paper, University of Chicago.*

DA COSTA, C. E., AND L. J. MAESTRI (2007): "The risk Properties of Human Capital and the Design of Government Policies," *European Economic Review*, 51(3), 695–713.

DIAMOND, P. (1998): “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates,” *American Economic Review*, 88(1), 83–95.

DIAMOND, P., AND J. SPINNEWIJN (2011): “Capital income taxes with heterogeneous discount rates,” *American economic journal: economic policy*, 3(4), 52–76.

DIAMOND, P. A., AND J. A. MIRRLEES (1978): “A model of Social Insurance with Variable Retirement,” *Journal of Public Economics*, 10(3), 295–336.

DIAMOND, P. A., AND E. SAEZ (2011): “The Case for a Progressive Tax: From Basic Research to Policy Recommendations,” *Journal of Economic Perspectives*, 25(4), 165–90.

DORNER, M., J. HEINING, P. JACOBEBBINGHAUS, AND S. SETH (2010): “The Sample of Integrated Labour Market Biographies,” *Schmollers Jahrbuch - Journal of Applied Social Science Studies*, 130(4), 599–608.

EATON, J., AND H. S. ROSEN (1980): “Taxation, human capital, and uncertainty,” *The American Economic Review*, 70(4), 705–715.

EBENSTEIN, A., A. HARRISON, M. MCMILLAN, AND S. PHILLIPS (2011): “Estimating the Impact of Trade and Offshoring on American Workers Using the Current Population Surveys,” Working Paper.

ELLISON, G., AND E. L. GLAESER (1997): “Geographic concentration in U.S. manufacturing industries: A dartboard approach,” *Journal of Political Economy*, 105(5), 889–927.

FAGGIO, G., AND H. G. OVERMAN (2012): “The Effect of Public Sector Employment on Local Labour Markets,” SERC Discussion Paper 111, London School of Economics.

FARHI, E., C. SLEET, I. WERNING, AND S. YELTEKIN (2012): “Non-linear Capital Taxation Without Commitment,” *The Review of Economic Studies*, 79(4), 1469–1493.

FARHI, E., AND I. WERNING (2008): “Optimal savings distortions with recursive preferences,” *Journal of Monetary Economics*, 55(1), 21–42.

FARHI, E., AND I. WERNING (2013): “Insurance and Taxation over the Life Cycle,” *The Review of Economic Studies*.

FEENSTRA, R. C., AND G. H. HANSON (1999): “The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the U.S., 1979-1990,” *Quarterly Journal of Economics*, 114(3), 907–940.

FEENSTRA, R. C., AND S.-J. WEI (2010): *China's Growing Role in World Trade*. NBER and University of Chicago Press.

FINDEISEN, S., AND D. SACHS (2012): "Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans," *Working Paper No. 40, Department of Economics, University of Zurich*.

FRIEDMAN, M. (1955): "The role of government in education," *Rutgers University Press*.

GALLIPOLI, G., C. MEGHIR, AND G. VIOLANTE (2011): "Equilibrium effects of education policies: A quantitative evaluation," *Working Paper*.

GELBER, A., AND M. WEINZIERL (2012): "Equalizing Outcomes vs. Equalizing Opportunities: Optimal Taxation when Children's Abilities Depend on Parents' Resources," *NBER Working Paper 18332*.

GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2011a): "Optimal Dynamic Taxes," *NBER Working Paper No. 17642.*

——— (2011b): "Optimal Dynamic Taxes," *NBER Working Paper No. 17642.*

GOLOSOV, M., M. TROSHKIN, A. TSYVINSKI, AND M. WEINZIERL (2012): "Preference heterogeneity and optimal capital income taxation," *Journal of Public Economics*.

GOLOSOV, M., AND A. TSYVINSKI (2006): "Designing optimal disability insurance: A case for asset testing," *Journal of Political Economy*, 114(2), 257–279.

GREENWOOD, J., Z. HERCOWITZ, AND G. W. HUFFMAN (1988a): "Investment, capacity utilization, and the real business cycle," *The American Economic Review*, pp. 402–417.

——— (1988b): "Investment, capacity utilization, and the real business cycle," *The American Economic Review*, pp. 402–417.

GROCHULSKI, B., AND T. PISKORSKI (2010): "Risky Human Capital and Deferred Capital Income Taxation," *Journal of Economic Theory*, 145(3), 908–943.

GROSSMAN, G. M., AND E. ROSSI-HANSBERG (2008): "Trading Tasks: A Simple Theory of Offshoring," *American Economic Review*, 98(5), 1978–1997.

GRUBER, J., AND E. SAEZ (2002): "The elasticity of taxable income: evidence and implications," *Journal of public Economics*, 84(1), 1–32.

HARRIGAN, J. (2000): "International Trade and American Wages in General Equilibrium, 1967-1995," in *The Impact of International Trade on Wages*, ed. by R. C. Feenstra, pp. 171 – 196. University of Chicago Press.

HARRISON, A., J. MCLAREN, AND M. MCMILLAN (2010): "Recent Findings on Trade and Inequality," NBER Working Paper No. 16425.

HEATHCOTE, J., F. PERRI, AND G. L. VIOLANTE (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic Dynamics*, 13(1), 15–51.

HENDRICKS, L., AND T. SCHOELLMAN (2012a): "Student Abilities During the Expansion of US Education," *Working Paper*.

——— (2012b): "Student Abilities During the Expansion of US Education, 1950-2000," .

HUGGETT, M., G. VENTURA, AND A. YARON (2011): "Sources of Lifetime Inequality," *American Economic Review*, 101.

JACOBS, B., AND A. BOVENBERG (2011a): "Optimal taxation of human capital and the earnings function," *Journal of Public Economic Theory*, 13(6), 957–971.

JACOBS, B., AND A. L. BOVENBERG (2011b): "Optimal Taxation of Human Capital and the Earnings Function," *Journal of Public Economic Theory*, 13(6), 957–971.

JACOBS, B., AND D. SCHINDLER (2012): "On the desirability of taxing capital income in optimal social insurance," *Journal of Public Economics*.

JACOBS, B., D. SCHINDLER, AND H. YANG (2012): "Optimal Taxation of Risky Human Capital," *Scandinavian Journal of Economics*, 114(3), 908–931.

JAPPELLI, T., AND L. PISTAFERRI (2010): "The Consumption Response to Income Changes," *Annu. Rev. Econ.*, 2(1), 479–506.

KAPICKA, M. (2006): "Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping," *Review of Economic Dynamics*, 9(4), 612–639.

——— (2013): "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach," *The Review of Economic Studies (forthcoming)*.

KAPICKA, M., AND J. NEIRA (2013): "Optimal taxation in a life-cycle economy with endogenous human capital formation," *Working paper, UC Santa Barbara.*

KAPLOW, L. (2006): "On the undesirability of commodity taxation even when income taxation is not optimal," *Journal of Public Economics*, 90(6), 1235–1250.

KARAHAN, F., AND S. OZKAN (2012): "On the persistence of income shocks over the life cycle: Evidence, theory, and implications," *Review of Economic Dynamics*.

KAUFMANN, D., A. KRAAY, AND M. MASTRUZZI (2010): "The worldwide governance indicators: methodology and analytical issues," *World Bank policy research working paper*, (5430).

KEANE, M., AND K. WOLPIN (1997): "The career decisions of young men," *Journal of Political Economy*, 105(3), 473–522.

KLEIBERGEN, F., AND R. PAAP (2006): "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 133(1), 97–126.

KLEVEN, H. J., AND E. A. SCHULTZ (2012): "Estimating Taxable Income Responses using Danish Tax Reforms," .

KONRAD, K. A. (2001): "Privacy and time-consistent optimal labor income taxation," *Journal of Public Economics*, 79(3), 503–519.

KOVAK, B. K. (2011): "Local Labor Market Effects of Trade Policy: Evidence from Brazilian Liberalization," mimeo, Carnegie Mellon University.

KROPP, P., AND B. SCHWENGLER (2011): "Abgrenzung von Arbeitsmarktregionen * ein Methodenvorschlag," *Raumforschung und Raumordnung*, 69(1), 45–62.

KRUEGER, D., AND A. LUDWIG (2013): "Optimal Progressive Taxation and Education Subsidies in a Model of Endogenous Human Capital Formation," *American Economic Review Papers & Proceedings (forthcoming)*.

KYDLAND, F. E., AND E. C. PRESCOTT (1977): "Rules rather than discretion: The inconsistency of optimal plans," *The Journal of Political Economy*, pp. 473–491.

LEMIEUX, T. (2006a): "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?," *The American Economic Review*, pp. 461–498.

——— (2006b): "Postsecondary Education and Increasing Wage Inequality," *American Economic Review*, 96(2), 195–199.

LUTTMER, E., AND R. ZECKHAUSER (2008): "Schedule selection by agents: from price plans to tax tables," *NBER Working Paper No. 13808*.

MANKIW, N., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23(4), 147–174.

MCLAREN, J., AND S. HAKOBYAN (2010): "Looking for Local Labor Market Effects of NAFTA," NBER Working Paper No. 16535.

MEGHIR, C., AND L. PISTAFERRI (2011): "Earnings, consumption and life cycle choices," *Handbook of Labor Economics*, 4, 773–854.

MENDOZA, E. G., AND V. Z. YUE (2012): "A general equilibrium model of sovereign default and business cycles," *The Quarterly Journal of Economics*, 127(2), 889–946.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38(2), 175–208.

MOLLOY, R., C. L. SMITH, AND A. K. WOZNIAK (2011): "Internal Migration in the United States," *Journal of Economic Perspectives*, 25(2), 1–42.

PAVAN, A., I. SEGAL, AND J. TOIKKA (2012): "Dynamic Mechanism Design," *Working Paper*.

PEREIRA, J. (2009): "Essays on Time-Consistent Fiscal Policy," Ph.D. thesis, European University Institute.

PIKETTY, T. (1997): "La redistribution fiscale face au chomage," *Revue française d'économie*, 12(1), 157–201.

PIKETTY, T., AND E. SAEZ (2013a): "Optimal labor income taxation," *Handbook of Public Economics, Vol 5.*

——— (2013b): "A Theory of Optimal Inheritance Taxation," *Econometrica*, forthcoming.

POLETAEV, M., AND C. ROBINSON (2008): "Human capital specificity: Evidence from the Dictionary of Occupational Titles and Displaced Worker Surveys, 1984-2000," *Journal of Labor Economics*, 26(3), 387–420.

RAMSAY, C. M. (2006): "The distribution of sums of certain iid Pareto variates," *Communications in Statistics- Theory and Methods*, 35(3), 395–405.

RICHTER, W. F. (2009): "Taxing education in Ramsey's tradition," *Journal of Public Economics*, 93(11), 1254–1260.

ROBERTSON, R. (2004): "Relative prices and wage inequality: evidence from Mexico," *Journal of International Economics*, 64(2), 387–409.

ROGERSON, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53(1), 69–76.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–229.

SAEZ, E. (2002): "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes," *Journal of Public Economics*, 83(2), 217–230.

SALANIÉ, B. (2003): *The Economics of Taxation*. The MIT Press.

SCHEUER, F. (2012): "Entrepreneurial Taxation with Endogenous Entry," *Working Paper, Stanford University*.

SHESHINSKI, E. (1972): "The optimal linear income-tax," *The Review of Economic Studies*, 39(3), 297–302.

SPENGLER, A. (2008): "The Establishment History Panel," *Schmollers Jahrbuch - Journal of Applied Social Science Studies*, 128(3), 501–509.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65(3), 557–586.

STANTCHEVA, S. (2012): "Optimal Taxation and Human Capital Policies Over the Lifecycle," *Graduate Student Research Paper 12-02 MIT*.

STANTCHEVA, S. (2013): "Optimal Income Taxation with Adverse Selection in the Labor Market," .

STIGLITZ, J. E. (1982): "Self-selection and Pareto efficient taxation," *Journal of Public Economics*, 17(2), 213–240.

STOCK, J. H., AND M. YOGO (2002): "Testing for Weak Instruments in Linear IV Regression," NBER Technical Working Paper No. 284.

STORESLETTEN, K., C. TELMER, AND A. YARON (2004): "Consumption and risk sharing over the life cycle," *Journal of Monetary Economics*, 51(3), 609–633.

SULLIVAN, D., AND T. VON WACHTER (2009): "Job Displacement and Mortality: An Analysis Using Administrative Data," *Quarterly Journal of Economics*, 124(3), 1265–1306.

TABER, C. (2001): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?," *Review of Economic Studies*, 68(3), 665–691.

TOPALOVA, P. (2010): "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," *American Economic Journal: Applied Economics*, 2(4), 1–41.

TOPEL, R. (1990): "Specific Capital and Unemployment: Measuring the Costs and Consequences of Job Loss," *Carnegie-Rochester Conference Series on Public Policy*, 33(1), 181–214.

TROSTEL, P. A. (1993): "The effect of taxation on human capital," *Journal of Political Economy*, pp. 327–350.

VERHOOGEN, E. (2008): "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *Quarterly Journal of Economics*, 123(2), 489–530.

VON WACHTER, T., AND S. BENDER (2006): "In the Right Place at the Wrong Time: The Role of Firms and Luck in Young Workers' Careers," *American Economic Review*, 96(5), 1679–1705.

WEINZIERL, M. (2011): "The Surprising Power of Age-Dependent Taxes," *The Review of Economic Studies*, 78(4), 1490–1518.

——— (2012): "Why Do We Redistribute So Much But Tag So Little? Normative Diversity and Optimal Taxation," *Working Paper, Harvard Business School*.

WERNING, I. (2007): "Optimal Fiscal Policy with Redistribution," *The Quarterly Journal of Economics*, 122(3), 925–967.

——— (2011): "Nonlinear Capital Taxation," *Working Paper, MIT*.

SEBASTIAN FINDEISEN

DATE AND PLACE OF BIRTH

June 10, 1983 in Mainz, Germany

ACADEMIC POSITIONS

2014 –	Assistant Professor	University of Mannheim
2013	Post-doc/Visiting Researcher	University of California at Berkeley

EDUCATION

2009 – 2013	Ph.D. Candidate in Economics	University of Zurich
2008 – 2009	MRes Economics	European University Institute, Florence
2003 – 2008	M.Sc., with Distinction, Economics	University of Konstanz

WORKING PAPERS

"Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans" (with Dominik Sachs)

"The Rise of the East and the Far East: German Labor Markets and Trade Integration" (with Wolfgang Dauth and Jens Suedekum)

"Education Policies and Taxation without Commitment" (with Dominik Sachs)

WORK IN PROGRESS

"Efficient Labor and Capital Income Taxation over the Life-Cycle" (with Dominik Sachs)

PUBLICATIONS

"Industry Churning and the Evolution of Cities: Evidence for Germany" (with Jens Suedekum), *Journal of Urban Economics*, 2008, Vol. 64 (2): 326-339.

PROFFESIONAL AFFILIATIONS

2010-	IZA, Research Affiliate

REFEREEING

Journal of the European Economic Association (4x), International Tax and Public Finance, German Economic Review

AWARDS AND FELLOWSHIPS

2013	Fellowship, Swiss National Science Foundation
2012	Reinhard-Selten-Award, Verein fuer Socialpolitik (German Economic Association)
2012-2013	Research Grant, University of Zurich
2011	Fellow (Handelsblatt) for the 4th Lindau Meeting of Nobel Laureates in Economic Sciences
2008	VEUK Prize for best Masters degree of the year in Economics, University of Konstanz
2008-2009	German Academic Exchange Service (DAAD) Fellowship, EUI
2007	Cash Prize, Graduate Student Paper Competition, North American Regional Science Council
2006-2008	Fellowship, University of Konstanz, Tuition Waiver
2005-2008	Fellowship, Friedrich-Ebert-Foundation
2005-2006	Fellowship, York University, Tuition Waiver